Herman Casier
Michiel Steyaert
Arthur H.M. van Roermund *Editors*

Analog Circuit Design

Robust Design, Sigma Delta Converters, RFID

Springer

VISIT…

Analog Circuit Design

Herman Casier • Michiel Steyaert
Arthur H. M. van Roermund
Editors

Analog Circuit Design

Robust Design, Sigma Delta Converters, RFID

Springer

Editors
Dr. Herman Casier
Consultant
Avondster 6
8520 Kuurne
Belgium
herman_casier@ieee.org

Michiel Steyaert
Dept. Elektrotechniek
ESAT-MICAS
K.U. Leuven
Kardinaal Mercierlaan 94
B-3001 HEVERLEE
Belgium
michiel.steyaert@esat.kuleuven.ac.be

Prof. Arthur H. M. van Roermund
Technical University Eindhoven
Electrical Engineering
Mixed-signal Microelectronics Group
Den Dolech 2
5600 MB Eindhoven
Netherlands
a.h.m.v.roermund@tue.nl

ISBN 978-94-007-0390-2 e-ISBN 978-94-007-0391-9
DOI 10.1007/978-94-007-0391-9
Springer Dordrecht Heidelberg London New York

© Springer Science+Business Media B.V. 2011
No part of this work may be reproduced, stored in a retrieval system, or transmitted in any form or by any means, electronic, mechanical, photocopying, microfilming, recording or otherwise, without written permission from the Publisher, with the exception of any material supplied specifically for the purpose of being entered and executed on a computer system, for exclusive use by the purchaser of the work.

Cover design: eStudio Calamar S.L.

Printed on acid-free paper

Springer is part of Springer Science+Business Media (www.springer.com)

Preface

This book is part of the *Analog Circuit Design* series and contains the revised contributions of all speakers of the 19th workshop on Advances in Analog Circuit Design (AACD), which was organized by Wolfgang Pribyl of Graz University of Technology. The workshop was held in the magnificent aula of the Graz University of Technology, Graz, Austria on March 23–25, 2010.

The book comprises 18 tutorial papers, divided in three chapters, each discussing a very relevant to-date topic in the area of analog circuit design. Each tutorial is presented by an expert in the field and state-of-the-art information is shared and discussed with the audience.

The topics of 2010 are:

1. **Robust Design**
2. **Sigma Delta Converters**
3. **RFID**

The aim of the AACD workshop is to bring together a group of expert designers to study and discuss new possibilities and future developments in the area of analog circuit design. Each AACD workshop has given rise to the publication of a book by Springer in their successful series of Analog Circuit Design. The series provides a valuable overview of analog circuit design and related CAD, mainly in the fields of basic analog modules, mixed-signal electronics, AD and DA converters, RF systems, robust and automotive electronics. It is a reference for whoever is engaged in these disciplines and wishes to keep abreast of the latest developments in the field. The full list of the previous books and topics in the series is enclosed below.

We sincerely hope that this 19th book continues the tradition and provides a valuable contribution to our Analog Design Community.

Herman Casier

Table Topics covered before in this series

2009	Lund (Sweden)	Smart Data Converters Filters on Chip Multimode Transmitters
2008	Pavia (Italy)	High-speed Clock and Data Recovery High-performance Amplifiers Power Management
2007	Oostende (Belgium)	Sensors, Actuators and Power Drivers for the Automotive and Industrial Environment Integrated PA's: from Wireline to RF Very High Frequency Front Ends
2006	Maastricht (The Netherlands)	High-Speed AD Converters Automotive Electronics: EMC issues Ultra Low Power Wireless
2005	Limerick (Ireland)	RF Circuits: Wide Band, Front-Ends, DAC's Design Methodology and Verification of RF and Mixed-Signal Systems Low Power and Low Voltage
2004	Montreux (Switzerland)	Sensor and Actuator Interface Electronics Integrated High-Voltage Electronics and Power Management Low-Power and High-Resolution ADCs
2003	Graz (Austria)	Fractional-N Synthesizers Design for Robustness Line and Bus drivers
2002	Spa (Belgium)	Structured Mixed-Mode Design Multi-Bit Sigma-Delta Converters Short-Range RF Circuits
2001	Noordwijk (The Netherlands)	Scalable Analog Circuit Design High-Speed D/A Converters RF Power Amplifiers
2000	Munich (Germany)	High-Speed A/D Converters Mixed-Signal Design PLLs and Synthesizers
1999	Nice (France)	(X)DSL and other Communication Systems RF-MOST Models and Behavioural Modeling Integrated Filters and Oscillators
1998	Copenhagen (Denmark)	1-Volt Electronics Mixed-Mode Systems LNAs and RF Power Amplifiers for Communications
1997	Como (Italy)	RF Analog to Digital Converters Sensor and Actuator Interfaces Low-Noise Oscillators, PLLs and Synthesizers
1996	Lausanne (Switzerland)	RF CMOS Circuit Design Bandpass Delta-Sigma and Other Data Converters Translinear Circuits
1995	Villach (Austria)	Low-Noise, Low-Power, Low-Voltage Mixed-Mode design with CAD tools Voltage, Current and Time References
1994	Eindhoven (The Netherlands)	Low-Power Low-Voltage Integrated Filters Smart Power

Table (continued)

1993	Leuven (Belgium)	Mixed Analogue-Digital Circuit Design Sensor Interface Circuits Communication Circuits
1992	Scheveningen (The Netherlands)	Operational Amplifiers Analog to Digital Conversion Analog Computer Aided Design

Contents

Part I Robust Design 1

Modeling and Design for Reliability of Analog Integrated Circuits in Nanometer CMOS Technologies 3
Georges Gielen, Elie Maricau and Pieter De Wit

Modeling and Simulation of Statistical Variability in Nanometer CMOS Technologies 17
A. Asenov and B. Cheng

Advanced Physical Design in Nanoscale Analog CMOS 35
Lanny L. Lewyn

Robust Design for High Temperature and High Voltage Applications 53
Ovidiu Vermesan, Edgard Laes, Marco Ottella, Mamun Jamal, Jan Kubik, Kafil M. Razeeb, Reiner John, Harald Gall, Massimo Abrate, Nicolas Cordero and Jan Vcelak

Radiation Effects and Hardening by Design in CMOS Technologies 69
Federico Faccio

EMC Robust Design for Smart Power High Side Switches 89
Paolo Del Croce and Bernd Deutschmann

Part II Sigma Delta Converters 105

Noise-Coupled Delta-Sigma ADCS 107
Kyehyung Lee and Gabor C. Temes

Very Low OSR Sigma-Delta Converters 135
Trevor C. Caldwell

Comparator-Based Switched-Capacitor Delta-Sigma A/D Converters 157
Koen Cornelissens and Michiel Steyaert

VCO-Based Wideband Continuous-Time Sigma-Delta Analog-to-Digital Converters 177
Michael H. Perrott

Wideband Continuous-Time Multi-Bit Delta-Sigma ADCs 203
J. Silva-Martinez, C.-Y. Lu, M. Onabajo, F. Silva-Rivas, V. Dhanasekaran and M. Gambhir

Oversampled DACs 227
Andrea Baschirotto, Vittorio Colonna and Gabriele Gandolfi

Part III RFID 257

RFID, a Technology Ready for Industry Deployment 259
Henri Barthel

The World's Smallest RFID Chip Technology 277
Mitsuo Usami

RF and Low Power Analog Design for RFID 289
Raymond Barnett

A Dual Frequency Band Comprehensive RFID TAG 313
Albert Missoni, Günter Hofer and Wolfgang Pribyl

Printed Electronics—First Circuits, Products, and Roadmap 333
Jürgen Krumm and Wolfgang Clemens

Towards EPC-Compatible Organic RFID Tags 347
Kris Myny, Soeren Steudel, Peter Vicca, Steve Smout, Monique J. Beenhakkers, Nick A. J. M. van Aerle, François Furthner, Bas van der Putten, Ashutosh K. Tripathi, Gerwin H. Gelinck, Jan Genoe, Wim Dehaene and Paul Heremans

Contributors

Massimo Abrate Centro Ricerche FIAT S.C.p.A., Strada Torino 50, 10043 Orbassano, Torino, Italy

Nick A. J. M. van Aerle Polymer Vision, Eindhoven, The Netherlands
ASML, Veldhoven, The Netherlands

A. Asenov Department of Electronics and Electrical Engineering, The University of Glasgow, Glasgow G12 0LT, UK

Raymond Barnett Texas Instruments, Texas, USA

Henri Barthel GS1 Global Office, Brussels, Belgium

Andrea Baschirotto Department of Physics "G. Occhialini", University of Milano Bicocca, Milano, Italy
e-mail: andrea.baschirotto@unimib.it

Monique J. Beenhakkers Polymer Vision, Eindhoven, The Netherlands

Trevor C. Caldwell Analog Devices, University of Toronto, Toronto, Canada

B. Cheng Department of Electronics and Electrical Engineering, The University of Glasgow, Glasgow G12 0LT, UK

Wolfgang Clemens PolyIC GmbH & Co.KG, Tucherstraße 2, 90763 Fürth, Germany

Vittorio Colonna Marvell, Pavia, Italy
e-mail: vcolonna@marvell.com

Nicolas Cordero Tyndall National Institute, Lee Maltings, Prospect Row, Cork, Ireland

Koen Cornelissens ESAT-MICAS, K.U. Leuven, Kasteelpark Arenberg 10, 3001 Leuven, Belgium

Paolo Del Croce Infineon Technologies AG, Siemensstrasse 2, 9500 Villach, Austria
e-mail: paolo.delcroce@infineon.com

Wim Dehaene IMEC, Leuven, Belgium;
Katholieke Universiteit Leuven, Leuven, Belgium

Bernd Deutschmann Infineon Technologies AG, Am Campeon 1–12, 85579 Neubiberg,Germany
e-mail: bernd.deutschmann@infineon.com

V. Dhanasekaran Electrical and Computer Engineering, Analog and Mixed-Signal Center, Texas A&M University, College Station, TX, USA

Federico Faccio PH dept., CERN, 1211 Geneva 23, Switzerland

François Furthner TNO Science and Industry, Eindhoven, The Netherlands

Harald Gall austriamicrosystems AG, Tobelbaderstrasse 30, 8141 Unterpremstaetten, Austria

M. Gambhir Electrical and Computer Engineering, Analog and Mixed-Signal Center, Texas A&M University, College Station, TX, USA

Gabriele Gandolfi Marvell, Pavia, Italy
e-mail: gabriele@marvell.com

Gerwin H. Gelinck TNO Science and Industry, Eindhoven, The Netherlands

Jan Genoe IMEC, Leuven, Belgium;
Katholieke Hogeschool Limburg, Diepenbeek, Belgium

Georges Gielen ESAT-MICAS, Katholieke Universiteit Leuven, Leuven, Belgium
e-mail: gielen@esat.kuleuven.be

Paul Heremans IMEC, Leuven, Belgium
Katholieke Universiteit Leuven, Leuven, Belgium

Günter Hofer Infineon Technologies Austria AG, Graz, Austria

Mamun Jamal Tyndall National Institute, Lee Maltings, Prospect Row, Cork, Ireland

Reiner John Infineon Technologies AG, Am Campeon 1-12, 85579 Neubiberg, Germany

Jürgen Krumm PolyIC GmbH & Co.KG, Tucherstraße 2, 90763 Fürth, Germany

Jan Kubik Tyndall National Institute, Lee Maltings, Prospect Row, Cork, Ireland

Edgard Laes ON Semiconductor Belgium BVBA, Senneberg, J. Monnetlaan, 1804 Vilvoorde, Belgium

Kyehyung Lee Conexant Systems, Newport Beach, CA 92660, USA

Lanny L. Lewyn Lewyn Consulting Inc., Laguna Beach, CA, USA
e-mail: lanny@pacbell.net

C.-Y. Lu Electrical and Computer Engineering, Analog and Mixed-Signal Center, Texas A&M University, College Station, TX, USA

Elie Maricau ESAT-MICAS, Katholieke Universiteit Leuven, Leuven, Belgium

Albert Missoni Infineon Technologies Austria AG, Graz, Austria;
Graz University of Technology Institute of Electronics, Graz, Austria

Kris Myny IMEC, Leuven, Belgium;
Katholieke Hogeschool Limburg, Diepenbeek, Belgium;
Katholieke Universiteit Leuven, Leuven, Belgium

Marco Ottella Centro Ricerche FIAT S.C.p.A., Strada Torino 50,
10043 Orbassano, Torino, Italy

M. Onabajo Electrical and Computer Engineering, Analog and Mixed-Signal Center, Texas A&M University, College Station, TX, USA

Michael H. Perrott SiTime Corporation, Sunnyvale, USA
e-mail: mhperrott@gmail.com

Wolfgang Pribyl Graz University of Technology Institute of Electronics,
Graz, Austria

Bas van der Putten TNO Science and Industry, Eindhoven, The Netherlands

J. Silva-Martinez Electrical and Computer Engineering, Analog and Mixed-Signal Center, Texas A&M University, College Station, TX, USA

Kafil M. Razeeb Tyndall National Institute, Lee Maltings, Prospect Row, Cork, Ireland

F. Silva-Rivas Electrical and Computer Engineering, Analog and Mixed-Signal Center, Texas A&M University, College Station, TX, USA

Steve Smout IMEC, Leuven, Belgium

Soeren Steudel IMEC, Leuven, Belgium

Michiel Steyaert Dept. Elektrotechniek, ESAT-MICAS, K.U. Leuven, Kardinaal Mercierlaan 94, B-3001 Heverlee, Belgium
e-mail: michiel.steyaert@esat.kuleuven.ac.be

Gabor C. Temes Oregon State University, Corvallis, OR 97331, USA

Ashutosh K. Tripathi TNO Science and Industry, Eindhoven, The Netherlands

Mitsuo Usami Central Research Laboratory, Hitachi Ltd., 1-280 Higashi-Koigakubo, Kokubunji-shi, 185-8601 Tokyo, Japan
e-mail: mitsuo.usami.fc@hitachi.com

Jan Vcelak Tyndall National Institute, Lee Maltings, Prospect Row, Cork, Ireland

Ovidiu Vermesan SINTEF, Forskningsvn. 1, P.O.Box 124 Blindern, 0314 Oslo, Norway

Peter Vicca IMEC, Leuven, Belgium

Pieter De Wit ESAT-MICAS, Katholieke Universiteit Leuven, Leuven, Belgium

Part I
Robust Design

In this first day of the AACD workshop in Graz, the state-of-the-art of robust design is discussed. The same topic also figured in the AACD program of 2003, also in Graz. The program then focussed on ESD in analog, smart power and RF applications, on EMC in automotive applications and on substrate coupling. The program of this year presents in-depth discussions of robust design in nanometer technologies, in high-temperature and high-voltage environments, in the presence of radiation and of conductive EMC.

The first three contributions deal with different aspects of robust design in nanometer technologies. The first addresses analysis tools and design solutions for the variability and reliability problems in analog circuits in nanometer CMOS technologies. A tool is presented to evaluate the impact of degradation on circuit performance and to identify reliability weak spots. It is also shown how digitally assisted analog design can be used to improve the resilience for variability and degradation of analog circuits.

The second contribution focuses on statistical variability in nanometer CMOS. Discreteness of charge and matter, atomic scale non-uniformities and material granularity are discussed and their expected impact on transistor parameters and on the variability sensitive SRAM cell in sub-45nm technologies is described. It is shown that a seven parameter statistical BSIM model accurately matches the physically simulated device characteristics.

The third paper identifies physical factors, which adversely affect sub-100nm CMOS analog designs and proposes methods for mitigating or overcoming them. It is shown that these methods require design restrictions at the circuit design stage as well as layout restrictions at the physical design stage. The historical separation between circuit and physical design activities must be replaced by a more close cooperation between them.

The next three papers deal with robust design for different hostile environments. The first of these contributions addresses high-temperature and high-voltage integrated electronics and modules for electric power systems in electric and hybrid-electric vehicles. Maximization of the compatibility between semiconductor, packaging and material technologies, thermal-electric simulation and reliability calculation of the whole configuration are key elements for robust smart power modules.

The next work is about radiation effects and hardening-by-design techniques in modern deep-submicron CMOS technologies. Both cumulative and single-event effects are discussed. Hardening-by-design techniques, leveraging on the intrinsic radiation tolerance of CMOS, are detailed for both effects. These techniques allow for the design of robust circuits with high radiation tolerance in a commercial-grade state-of-the-art CMOS technology instead of a dedicated radiation-hard technology.

The last contribution first shows how the standardized direct power injection measurement test bench can be transferred to the simulation environment to predict the EMC immunity of a smart power circuit. This simulation environment is then used to optimize a new methodology for robust high-side smart power switches, based on pin impedance control. Measurements endorse the validity and accuracy of this approach.

Herman Casier

Modeling and Design for Reliability of Analog Integrated Circuits in Nanometer CMOS Technologies

Georges Gielen, Elie Maricau and Pieter De Wit

1 Introduction

The evolution towards nanometer CMOS technologies (90, 65, 45 nm and below) [1] has enabled the design of complex Systems on a Chip (SoC) in consumer-market applications such as telecom and multimedia. These integrated systems are increasingly mixed-signal designs, embedding high-performance analog or mixed-signal blocks and possibly sensitive RF frontends together with complex digital circuitry (multiple processors, some logic blocks, and several large memory blocks) on the same chip. Even when developing a heterogeneous System in Package (SiP), the digital dies will likely have several embedded analog blocks.

The use of CMOS nanometer technologies however also brings significant challenges for the actual circuit design (both analog and digital) that were not encountered before. These challenges include [2]:

- managing the ever increasing design complexities in tightening time-to-market constraints, which requires proper EDA methodologies and tools to increase designer efficiency (e.g. using analog synthesis tools), as well as to verify full mixed-signal systems;
- the increasing variability of technology parameters, causing mismatch and yield problems;
- aggravating degradation mechanisms (e.g. NBTI, Hot Carriers) and increasing reliability constraints such as EMC/EMI regulations.

This chapter addresses analysis tools and design solutions for the variability and reliability problems in analog circuits. After describing the degradation phenomena, it presents tools for the efficient analysis and identification of reliability problems in analog circuits. In addition, it describes circuit techniques for run-time reconfiguration of the analog circuits to make them degradation-resilient. The chapter is organized as follows. Section 2 briefly describes the problems of variability- and

G. Gielen (✉)
ESAT-MICAS, Katholieke Universiteit Leuven, Leuven, Belgium
e-mail: gielen@esat.kuleuven.be

H. Casier et al. (eds.), *Analog Circuit Design,*
DOI 10.1007/978-94-007-0391-9_1, © Springer Science+Business Media B.V. 2011

reliability-induced circuit performance degradation. Section 3 outlines an approach for the reliability analysis and the detection of reliability weak spots in analog circuits. An efficient extension to include the effect of variability is included as well. Section 4 then presents run-time circuit self-adaptation techniques to make them degradation-resilient. This concept is illustrated with a practical design example. Finally, Sect. 5 provides conclusions.

2 Variability- and Reliability-Induced Analog Circuit Performance Degradation

Many non-idealities in analog circuits originate from random and systematic errors in the implementation of the circuit. These errors represent the time-independent reliability problems in a circuit. Although the random errors are unknown at the time of design, they can largely be compensated or calibrated away after fabrication. Time-dependent degradation on the other hand occurs due to ageing of the transistors over time, which results in the circuit performance changing over time, hence possibly turning initially fully functional circuits into nonfunctional circuits. Also disturbance sources such as electromagnetic interference (EMI) or substrate noise coupling may turn functional analog circuits into nonfunctional circuits, but normally don't change performance permanently, so we will not focus on these in this chapter. We will now describe the variability and degradation effects in some more detail.

2.1 Variability and Mismatch

Random errors, usually denoted as variability, are the result of the stochastic nature of many physical processes that take place during the fabrication of integrated circuits, such as line edge roughness and random dopant fluctuations. In analog circuits device mismatch between identically designed devices is a key limitation to the accuracy of the circuit. For example, to first order the threshold voltage mismatch ΔV_T between identical transistors with active area WL is given by [3, 4]:

$$\sigma^2(\Delta V_T) = \frac{A_{VT}^2}{WL} + S_{VT}^2 D^2 \tag{1}$$

where A_{VT} and S_{VT} are process-dependent constants and D is the distance between the devices. Unfortunately, as can be seen from (1), this mismatch is inversely proportional to the area of the devices and therefore can only be improved by making the devices larger. This implies more chip area and more power for the same speed.

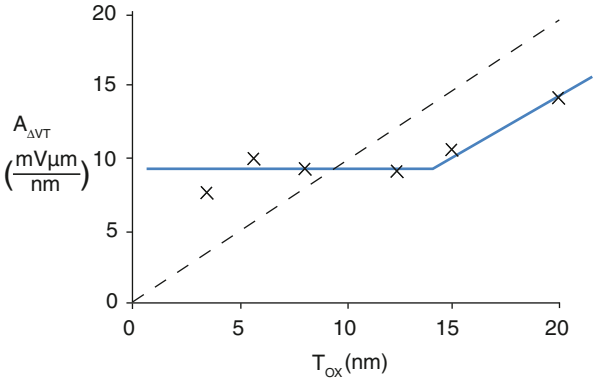


Fig. 1 Evolution of the threshold voltage mismatch parameter as a function of the gate oxide thickness according to [5]. ($A_{\Delta VT}$ is the parameter characterizing threshold voltage mismatch in (1) [3, 4])

In addition, as shown in Fig. 1, Tuinhout observed that, although historically this mismatch was improving with scaling technology, for values of gate oxide thickness below 10 nm, this improvement seems to level off, indicating that mismatch is no longer improving much with technology progress. This might significantly reduce the benefits of technology scaling for analog integrated circuits [5].

2.2 Time-Dependent Degradation

Time-dependent degradation effects cause a change of the transistor parameters (V_T, β, r_o) as a function of time and therefore might turn an initially fully functional circuit into a less or even non-functional circuit over time [6]. This degradation depends on the stress applied to the device, i.e. the voltages and currents applied to the transistor, as well as the transistor sizes and the temperature. Figure 2 qualitatively indicates the impact of these mechanisms on the I_{DS}–V_{DS} characteristic of a MOS device for an arbitrary stress time. Some of the most important degradation mechanisms will now be discussed in more detail.

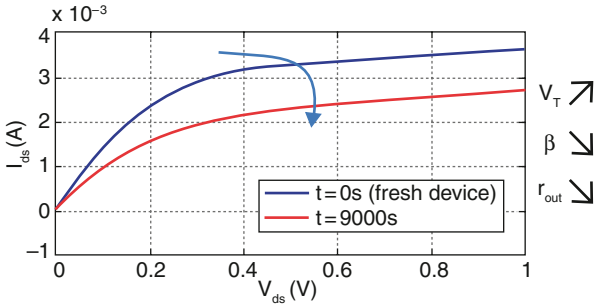


Fig. 2 Time-dependent variation of the characteristics of a transistor due to Hot-Carrier and NBTI degradation [6]

2.2.1 Time-Dependent Dielectric Breakdown

In ultra-scaled MOSFETs, the strong electric fields across the gate oxide can cause oxide damage leading to dielectric breakdown (BD), i.e. the loss of the isolating properties of the oxide. BD is an extremely local phenomenon, for which an extra current flows through a small region of the gate oxide. Prior to oxide BD, a degradation process of the dielectric takes place that initiates the generation of traps in random positions inside the oxide and at the interface. A stress-induced leakage current is produced during this degradation stage. If the dielectric degradation increases, a critical trap density is reached and BD occurs [7, 8]. Due to this behavior the time to BD can be described using a Weibull probability distribution. Obviously this breakdown phenomenon influences the circuit performance over time, eventually resulting in circuit failure.

2.2.2 Hot-Carrier Injection

A second degradation phenomenon is Hot Carrier Injection (HCI), which manifests itself mainly as a threshold voltage shift, and some degradation of carrier mobility and a change of output resistance [9, 10]. During hot carrier stress, which consists of a large electric field near the drain end of a transistor in saturation, hot carriers are produced. These carriers introduce both oxide and interface traps (near the drain) and a substrate current. As holes are much 'cooler' than electrons, hot carrier effects in nMOS devices are proven to be more significant than in pMOS devices [11]. Removal of the stress anneals some of the interface traps, resulting in partial recovery. But as these traps are only present at the drain junction of the transistor, this recovery is negligible in comparison to NBTI relaxation (see next subsection).

HCI degradation is typically modeled with a power law dependence on the stress time t [10]. For example, the increase in threshold voltage V_T over time can be written as:

$$\Delta V_T = A(V_{GS}, V_{DS}, T, W, L \ldots)\, t^n \tag{2}$$

where the function A depends on the applied voltages, the temperature T but also on the transistor width W and length L. In the case of HCI, the exponent n is in the order of 0.45. A full model can be found in [12]. Due to hot carrier degradation, transistors degrade over time, which translates to a change in the circuit performance over time. Therefore, tools need to be developed to analyze this problem and to identify possible reliability problems in a circuit before tapeout. This will be described in Sect. 3. Circuit techniques to deal with this kind of degradation also need to be developed and will be described in Sect. 4.

2.2.3 Negative Bias Temperature Instability

Negative Bias Temperature Instability (NBTI) has an increasingly adverse impact on nanometer CMOS technology [13]. NBTI is typically seen as a threshold voltage

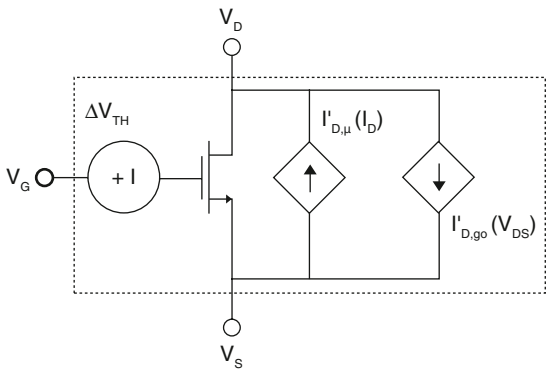


Fig. 3 Equivalent model of transistor to model degradation due to Hot Carriers or NBTI

shift after a negative bias has been applied to a MOS gate at elevated temperature, hence mainly affecting pMOS transistors [14]. Degradation of channel carrier mobility is also observed. The NBTI degradation is typically represented as following a power law with stress time t similar to Eq. 2, but with the exponent n typically being around 0.18 and with a different function A. The reciprocal PBTI behavior in nMOS transistors is still less important.

A peculiar property of the NBTI mechanism is the so-called relaxation or recovery of the degradation immediately after the stress voltage has been reduced [15]. This relaxation behavior greatly complicates the evaluation of NBTI, its modeling and the extrapolation of its impact on circuitry. A complete model of NBTI for circuit analysis has however not yet been fully established.

Since HCI and NBTI impact both the transistor parameters V_T, β, *and* r_o, this can be modeled by replacing the transistor in the circuit netlist by a subcircuit as shown in Fig. 3. The additional sources model the change in V_T, β, *and* r_o according to Eq. 2, so that the intrinsic device model, e.g. BSIM, can be used for the actual transistor. This allows to simulate a degraded circuit with a standard SPICE simulator, with an additional script to update the values of the extra sources according to Eq. 2.

3 Reliability Analysis of Analog Integrated Circuits

Due to the impact of degradation on integrated circuits, it is important to analyze quantitatively the impact of HCI and NBTI on the performance and lifetime of a circuit, and to identify potential reliability problems in a circuit at design time, so that—if needed—the design can be modified to guarantee correct functionality and performance over the lifetime of the electronic product. Therefore, proper reliability simulation and analysis tools are needed for analog circuits.

In [16] an efficient method for reliability simulation of electronic circuit is presented. It uses a short transient simulation that provides accurate information about the stress at every circuit node, while a degradation extrapolation ensures a fast simulation result. Figure 4 gives a schematic representation of this reliability simulation algorithm. The input to the simulator is a fresh (i.e. unstressed) netlist. A transient simulation over time T_{tr} with step size t_{tr} is performed on the input netlist. As circuit

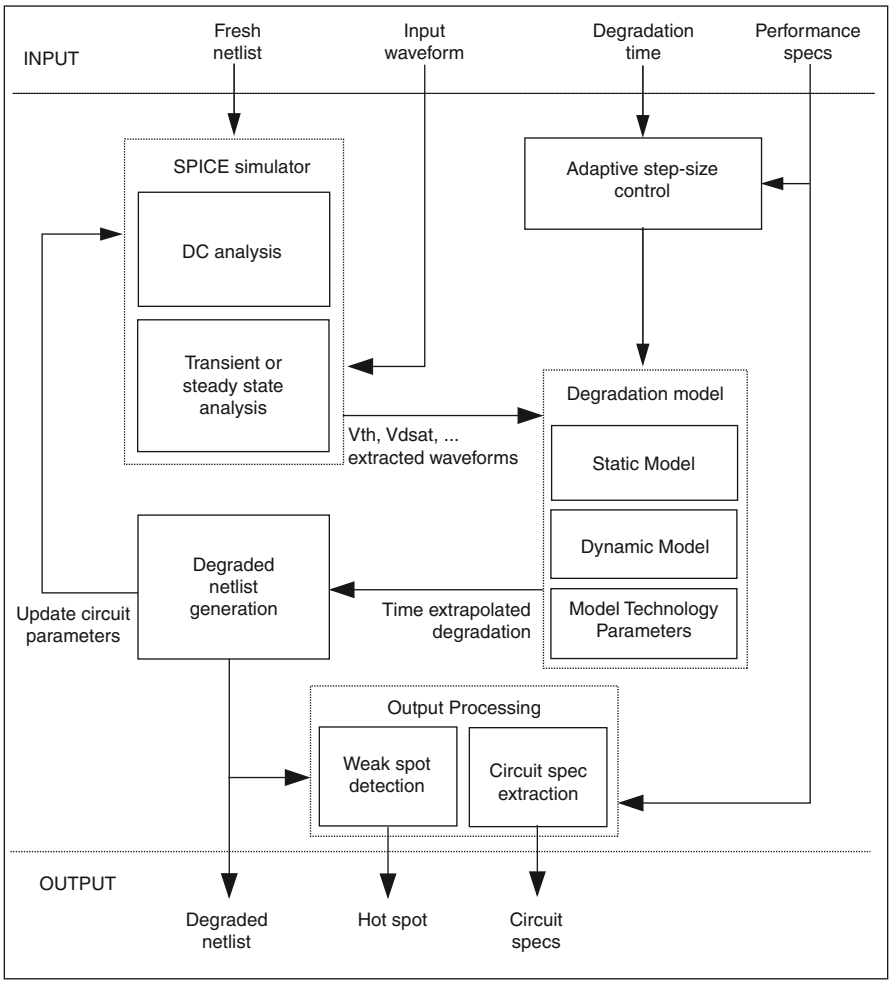


Fig. 4 Flow diagram of efficient degradation analysis of analog circuits [16]

input a periodic time-varying signal, $v_{in}(t)$, with period T_{in} is applied. The impact of completely arbitrary (i.e. non-periodic) input signals can only be calculated using a transient analysis over the entire operating period of the circuit, which obviously is not very feasible. Since the operation of most electronic systems is based on the execution of periodically executed algorithms (e.g. MPEG encoding), the restriction to periodic input signals is not a real limitation. Once the stress pattern on every transistor node is calculated from this initial simulation, it is extracted and passed on to a degradation model based on Eq. 2, which extrapolates the transistor degradation to the desired circuit operation time. Finally, a degraded version of the netlist is created as an output, using equivalent models like Fig. 3 for the individual transistors. A designer can use this output netlist to study the impact of degradation on the circuit performances and to identify the reliability weak spots.

In [17] the simulation method has also been extended to take variability of the device parameters into account. The initial approach of using Monte-Carlo simulations was replaced by design of experiments (DOE) techniques, resulting in a speed improvement of up to 4 orders of magnitude compared to standard Monte-Carlo, while maintaining the same simulation accuracy. In order to obtain a quasi-linear computational complexity, first a screening is performed to eliminate unimportant parameters, followed by a set of regression designs to build a response surface model used for the actual statistical analysis.

3.1 Illustrative Circuit Example

To illustrate the variability-aware reliability simulation methodology, the LC-VCO of Fig. 5 is analyzed [17]. Both HC and NBTI degradation are included. The example is simulated in a 90 nm CMOS technology. Large voltages at the drains of the cross-coupled transistor pair cause these transistors to suffer from hot-carrier degradation. Figure 6 shows the effect of degradation on both the oscillation frequency and the output swing for nominal design parameters (no statistical variations). The

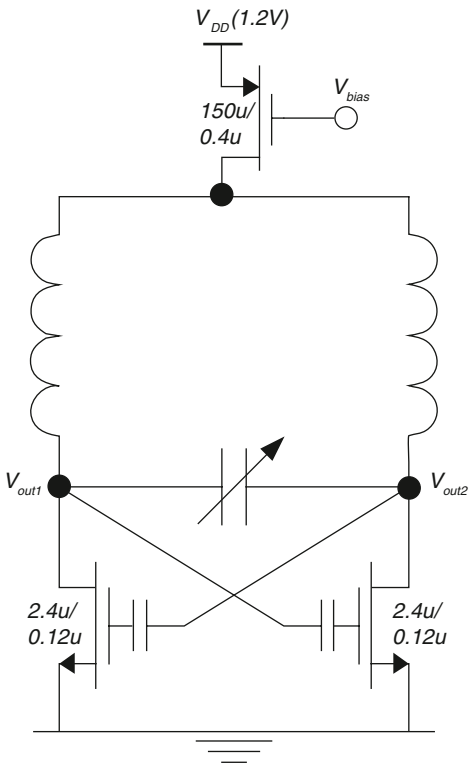


Fig. 5 Circuit schematic of the LC-VCO

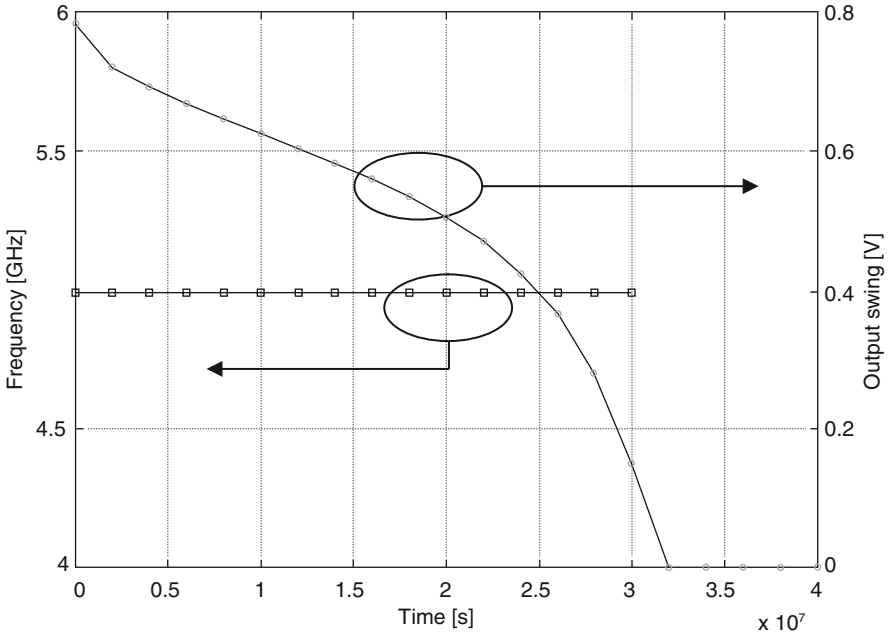


Fig. 6 Frequency and output swing of the LC-VCO as a function of time due to hot carrier degradation

frequency of the LC-oscillator is not affected by degradation; the output swing, however, decreases significantly over time. For the (slightly overstressed) settings used, the circuit fails after six months, if we assume that the device is considered to be defective when Vout < 0.6 V.

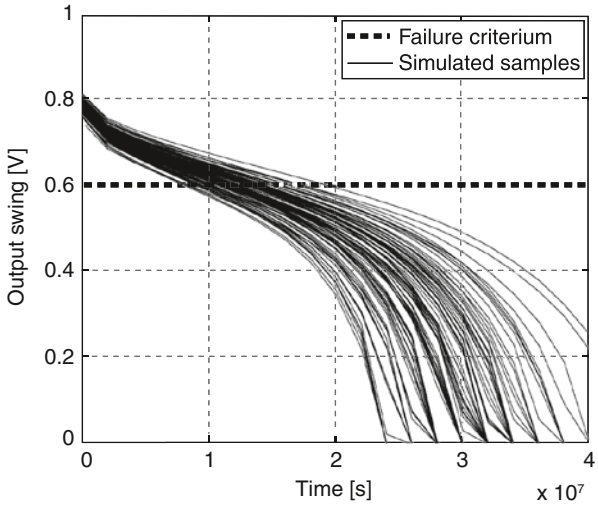


Fig. 7 Variability-induced dispersion of the time to failure of the LC-VCO

Now let us consider process variability. Figure 7 shows the failure time dispersion of the oscillator due to local and global errors of transistor parameters. Clearly, there is a large spread. Although a reliability simulation of the nominal circuit (not including variability) indicates a failure time of six months (see Fig. 6), running a variability-aware reliability simulation shows that 20% of the circuits already fail in less than four months.

4 Circuit Techniques for Variability and Degradation Resilience

The mechanisms described in Sect. 2 may cause serious performance changes and hence reliability problems in nanometer CMOS electronic systems. The classical approaches, i.e. guaranteeing intrinsic robustness by worst-case overdesign or use of redundancy, introduce an unacceptable power and area penalty. In order to obtain a high-performance and reliable system, using nanometer technologies, novel circuit design techniques are needed to deal with variability and time-dependent degradation. Some of these will be discussed here briefly.

4.1 Solutions to Processing Variability

Although mismatch is statistical in the sense that it is unknown before fabrication, it is fixed after fabrication. Therefore, all such static time-independent errors can to large extent be compensated for after fabrication using post-fabrication calibration methods. An example of this calibration technique has been proposed and verified on silicon in [18] where it is applied on the design of a 14-bit 200 MHz current-steering DAC. The high accuracy is obtained using a Switching-Sequence Post-Adjustment (SSPA) calibration technique, which dynamically rearranges the switching sequence of the unary MSB current sources. Since this technique is applied after chip fabrication (i.e. calibration at run time), random errors can partially be cancelled out. As a result, the area requirement imposed by the *INL* property ($INL < 0.5LSB$), which intrinsically would require a large area to sufficiently reduce the mismatch, is reduced dramatically to only 6% of the area of an intrinsic-accuracy DAC [19]. The only extra analog building block needed is an accurate current comparator. A chip photograph is shown in Fig. 8. The total area of the chip is 3 mm^2, where the area of the analog part is only 0.28 mm^2. The area of the digital part (mainly the calibration controller) is still quite large in this version, but will be scaled significantly in advanced processes.

This type of heavy-digital analog circuit design corresponds to the paradigm of digitally assisted analog circuits [20], where the analog part is actually under-sized to meet the required performances intrinsically, but where digital circuitry corrects for the errors or the nonidealities through reconfiguration or calibration.

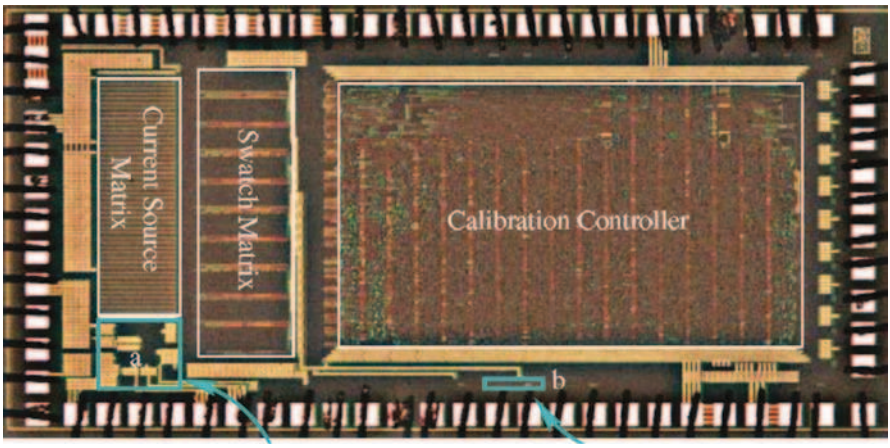


Fig. 8 Chip microphotograph of current-steering DAC with post-fabrication calibration [18]

4.2 Solutions to Time-Dependent Degradation

The solution to the time-dependent problem can be found in the use of monitors and knobs as presented in [21, 22]. The idea is to continuously monitor the operation of a system or circuit and take runtime countermeasures to compensate for variability and reliability errors. This guarantees a correct and optimal operation at all times, if properly anticipated at design time by using analysis and design tools such as presented in Sect. 3. As shown in Fig. 9, such a system consists of three parts. *Monitors* measure the actual performance of the system. Simple measurement circuits are required to achieve this. *Knobs* are tunable or reconfigurable circuit parts able to change the operating settings of the system. Finally, a *Control Algorithm* selects, based on the inputs from the different monitors, the optimal configuration of the system knobs in order to satisfy the system specifications, even if the performance varies over time. The control loop can be implemented in digital hardware, adding only a limited amount of extra power dissipation and area consumption. Note that this concept can be applied to both analog and digital circuits.

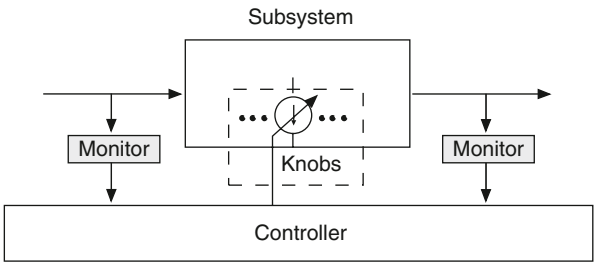


Fig. 9 Principle schematic of knobs and monitors concept to overcome degradation and failures

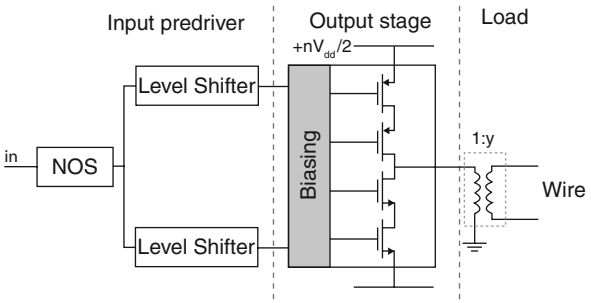


Fig. 10 Schematic of high-voltage line driver

The advantages of a system with knobs and monitors are:

- a self-adaptive system is obtained. It compensates for variability and degradation-induced errors as they occur, thus keeping the system in its optimal operation point of specifications and operating conditions.
- overdesign is not needed anymore. Design specifications for a given circuit can be relaxed because multiple possible operating modes exist within one system. This results in an overall decrease in power dissipation and area consumption of the entire system. Switching to another operating point might cause a slightly larger power consumption, but correct operation is guaranteed at any time, which is a reasonable trade-off in safety-critical applications that demand absolute reliability.

As an example, Fig. 10 shows a high-voltage line driver circuit. Due to the high voltages used, the output transistors degrade over time in nanometer CMOS technologies. Since the power efficiency of the line driver is directly linked to the on-resistance of the output transistors which degrades over time, solutions to guarantee the reliable operation with high power efficiency have to be found. As shown in Fig. 11, this can be solved by making the circuit reconfigurable [23]. Extra sub-transistor blocks P_i and N_i are added, besides the monitor circuits and the controller. The resilience of the circuit is defined as the minimum efficiency loss the system can detect and correct. If the circuit exceeds this limit, i.e. if the efficiency drops below the minimum set value, then the circuit needs to reconfigure and heal itself to "crank up" the efficiency again. To this end, extra sub-transistors (the extra "blocks" in Fig. 11) are added to the output transistors. These sub-transistors can be switched on under digital control. Monitor circuits are added to the line driver to monitor the actual degradation of the output transistors. They feed their information to the controller that decides on turning on extra sub-transistors.

Obviously, all this requires extra chip area. The system-level trade-off between the granularity of the extra sub-transistors, the circuit resilience and the extra area overhead has been elaborated in [23]. Figure 12 shows the area overhead as a function of the ripple on the power efficiency $\Delta\eta$. Clearly, for a degradation of

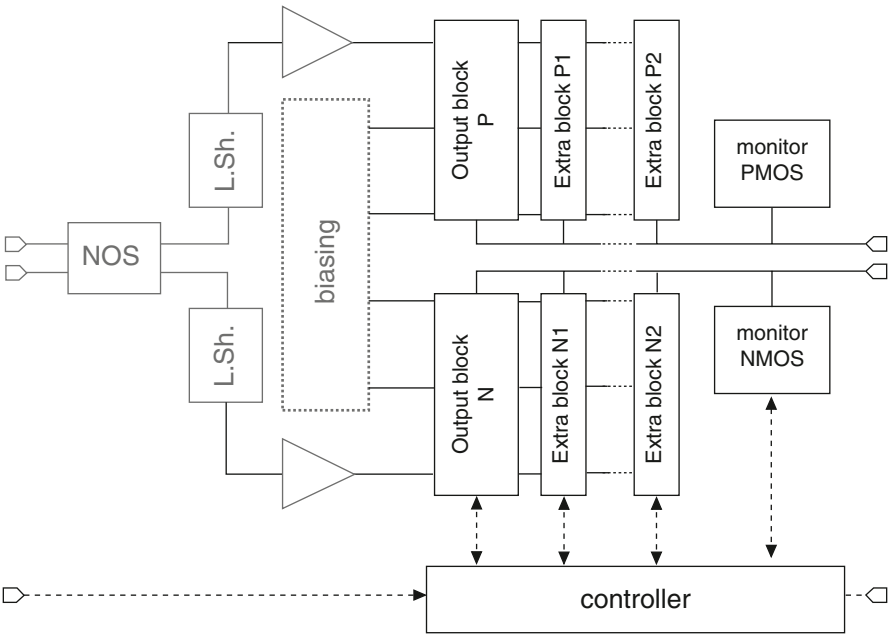


Fig. 11 Modified output stage of the high-voltage line driver with extra parallel sub-transistors to make the circuit failure-resilient [23]

the power efficiency with maximum a couple percent, the circuit can be made degradation-resilient with an area overhead of 4–5%, which is quite acceptable compared to worst-case overdesign of the circuit. Also the power consumption is improved.

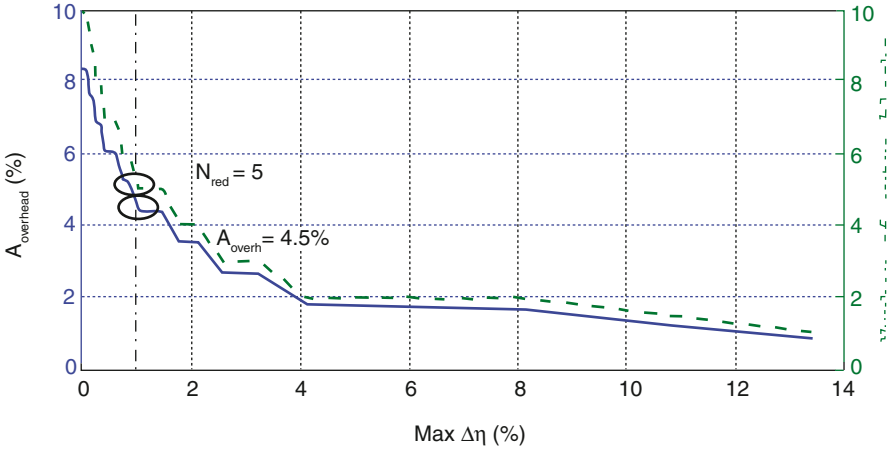


Fig. 12 Trade-off between area overhead and granularity of the extra subtransistors, which is directly related to the allowed "ripple" $\Delta\eta$ on the power efficiency [23]

5 Conclusions

CMOS technology is evolving deeper and deeper into the nanometer era, enabling the integration of entire systems, many of which are mixed-signal in nature. This paper has described solutions to variability and reliability challenges posed by these nanometer CMOS technologies. The problem of time-dependent performance degradation due to breakdown, hot carriers or NBTI has been described and models have been proposed. Next, an efficient tool for the analysis and identification of reliability problems in analog circuits has been described. The tool allows identifying potential ageing problems in circuits. An efficient approach to analyze the effect of variability on the reliability has been developed too. Finally, run-time circuit adaptation/reconfiguration techniques have been proposed to make circuits failure-resilient so that they can self-recover from degradation. These techniques are fully compliant with the trend towards digitally assisted analog circuits. All this has been illustrated with several design examples.

Acknowledgements The authors acknowledge the financial support of FWO, IWT SBO Elixir and EU FP7 Reality for different parts of this work.

References

1. ITRS 2008 technology roadmap. http://public.itrs.net/
2. G. Gielen, W. Dehaene, Analog and digital circuit design in 65 nm CMOS: end of the road?, in *Proceedings DATE Conference*, March 2005, vol. 1, pp. 37–42
3. K. Lakshmikumar et al., Characterisation and modeling of mismatch in MOS transistors for precision analog design. IEEE J. Solid-State Circ. **21**(6), 1057–1066 (1986)
4. M. Pelgrom et al., Matching properties of MOS transistors. IEEE J. Solid-State Circ. **24**(5), 1433–1439 (1989)
5. H. Tuinhout, Impact of parametric mismatch and fluctuations on performance and yield of deep-submicron CMOS technologies, in *Proceedings ESSDERC*, 2002
6. G. Gielen et al., Emerging yield and reliability challenges in nanometer CMOS technologies, in *Proceedings DATE Conference*, 2008, pp. 1322–1327
7. J. Stathis, Physical and predictive models of ultrathin oxide reliability in CMOS devices and circuits. IEEE Trans. Dev. Mater. Reliab. **1**, 43 (2001)
8. B. Kaczer et al., Analysis and modeling of a digital CMOS circuit operation and reliability after gate oxide breakdown: a case study. Microelectron. Reliab. **42**, 555 (2002)
9. I. Kurachi et al., Physical model of drain conductance, gd, degradation of NMOSFET's due to interface state generation by hot carrier injection. IEEE Trans. Electron. Dev. **41**, 1618 (1994)
10. W. Wang et al., Compact modeling and simulation of circuit reliability for 65 nm CMOS technology. IEEE Trans. Dev. Mater. Reliab. **7**(4) 509–517 (2007)
11. C. Hu et al., Hot-electron-induced MOSFET degradation—model, monitor, and improvement. IEEE Trans. Electron. Dev. **ED-32**, 375–384 (1985)
12. E. Maricau, P. De Wit, G. Gielen, An analytical model for hot carrier degradation in nanoscale CMOS suitable for the simulation of degradation in analog IC applications. Microelectron. Reliab. **48**(8–9), 1576–1580 (2008)
13. D. Schroder et al., Negative bias temperature instability: road to cross in deep submicron silicon semiconductor manufacturing. J. Appl. Phys. **94**, 1 (2003)

14. J. Stathis et al., The negative bias temperature instability in MOS devices: a review. Microelectron. Reliab. **46**, 270–286 (2006)
15. G. Chen et al., Dynamic NBTI of PMOS transistors and its impact on device lifetime, in *Proceedings Reliability Physics Symposium*, 2003
16. E. Maricau, G. Gielen, Reliability simulation of analog ICs under time-varying stress in nanoscale CMOS, in *Proceedings DRV Workshop,* 2008
17. E. Maricau, G. Gielen, A quasi-linear deterministic variation-aware circuit reliability simulation methodology, in *Proceedings DATE Conference,* 2010
18. T. Chen, G. Gielen, A 14-bit 200-MHz current-steering DAC with switching-sequence post-adjustment calibration. IEEE J. Solid-State Circ. **42**, 2386–2394 (2007)
19. G. Van der Plas, et al., A 14-bit intrinsic accuracy Q^2 random walk high-speed CMOS DAC. IEEE J. Solid-State Circ. **34**(12), 1708–1718 (1999)
20. B. Murmann, Digitally assisted analog circuits. IEEE Micro. **26**(2), 38–47 (2006)
21. A. Papanikolaou, Reliability issues in deep deep submicron technologies: time-dependent variability and its impact on embedded system design. 13th IEEE International On-Line Testing Symposium (IOLTS 2007)
22. B. Dierickx, Designing with unreliable components, invited presentation at the 2007 IET-FSA meeting, 14-15 May 2007, Paris
23. P. De Wit, G. Gielen, System-level design of a self-healing, reconfigurable output driver, in *Proceedings DRV Workshop,* 2008

Modeling and Simulation of Statistical Variability in Nanometer CMOS Technologies

A. Asenov and B. Cheng

1 Introduction

The years of 'happy scaling' are over and the fundamental challenges that the semiconductor industry faces at technology and device level will deeply affect the design of the next generations of integrated circuits and systems. The progressive scaling of CMOS transistors to achieve faster devices and higher circuit density has fuelled the phenomenal success of the semiconductor industry—captured by Moore's famous law [1]. Silicon technology has entered the nano CMOS era with 35 nm MOSFETs in mass production in the 45 nm technology generation. However, it is widely recognised that the increasing variability in the device characteristics is among the major challenges to scaling and integration for the present and next generation of nano CMOS transistors and circuits. The statistical variability of transistor characteristics, which has been previously concern only in the analogue design domain, has become a major concern associated with CMOS transistors scaling and integration [2, 3]. It already critically affects SRAM scaling [4], and introduces leakage and timing issues in digital logic circuits [5].

In the next section we review the major sources of statistical variability in nano CMOS transistors focusing at the 45 nm technology generation and beyond. In Sect. 3 we use advanced 3D physical statistical simulation to forecasts the magnitude of statistical variability in contemporary and future CMOS devices. The compact model strategies suitable for capturing the statistical variability in industrial strength compact models such as BSIM and PSP are outlined in Sect. 4. Section 5 presents example of statistical SRAM circuit simulation employing the statistical compact model strategies discussed in Sect. 4. Finally the conclusions are drawn in Sect. 6.

A. Asenov (✉)
Department of Electronics and Electrical Engineering, The University of Glasgow, Glasgow, G12 0LT, UK

H. Casier et al. (eds.), *Analog Circuit Design,*
DOI 10.1007/978-94-007-0391-9_2, © Springer Science+Business Media B.V. 2011

2 Sources of Statistical Variability

The statistical variability in modern CMOS transistors is introduced by the inevitable discreteness of charge and matter, the atomic scale non-uniformity of the interfaces and the granularity of the materials used in the fabrication of integrated circuits.

The granularity introduces significant variability when the characteristic size of the grains and irregularities become comparable to the transistor dimensions. For conventional bulk MOSFETs, which are still the workhorse of the CMOS technology, Random Discrete Dopants (RDD) are the main source of statistical variability [6]. Random dopants are introduced predominantly by ion implantation and redistributed during high temperature annealing. Figure 1 illustrates the dopant distribution obtained by the atomistic process simulator DADOS by Synopsys. Apart from special correlation in the dopant distribution imposed by the silicon crystal lattice, there may be also correlations introduced by the Coulomb interactions during the diffusion process. Line Edge Roughness (LER) illustrated in Fig. 2 stems from the molecular structure of the photoresist and the corpuscular nature of light. The polymer chemistry of the 193 nm lithography used now for few technology generations mainly determines the current LER limit of approx 5 nm [7]. In transistors with poly-silicon gate Poly Gate Granularity (PGG) illustrated in Fig. 3 is another important source of variability. This is associated surface potential pinning at the grain boundaries complimented by doping non-uniformity due to rapid diffusion along the grain boundaries [3].

The introduction of high-k/metal gate technology improves the RDD induced variability, which is inversely proportional to the equivalent oxide thickness (EOT). This is due to the elimination of the polysilicon depletion region and better screening of the RDD induced potential fluctuations in the channel from the very high concentration of mobile carriers in the gate. The metal gate also eliminates the PGG induced variability. In the same time it introduces high-k granularity illustrated in Fig. 4 and variability due to work-function variation associated with the metal gate granularity

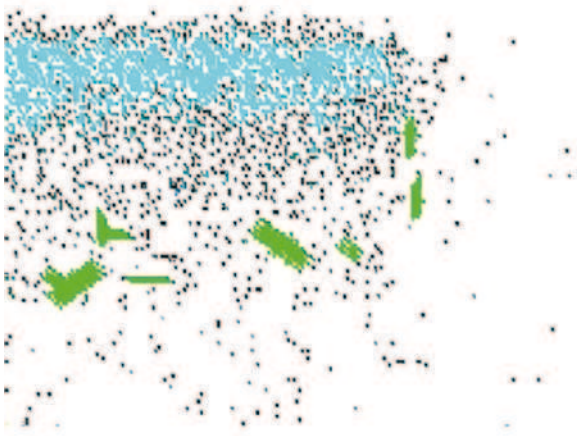


Fig. 1 KMC simulation of RDD. (DADOS, Synopsys)

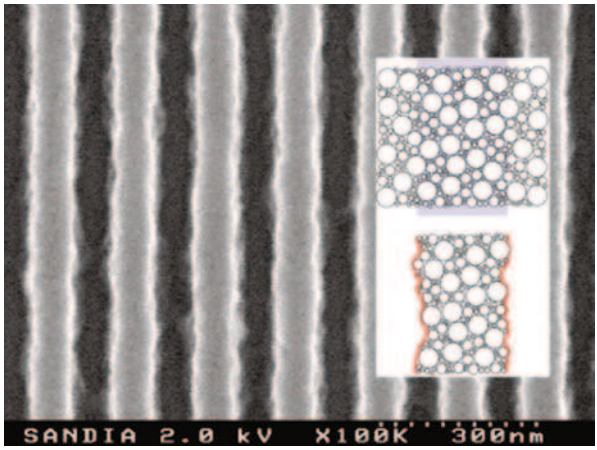


Fig. 2 Typical LER in photoresist. (Sandia Labs)

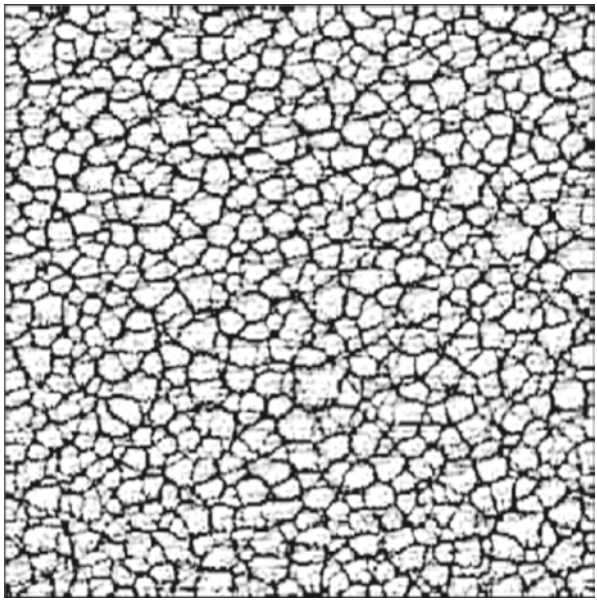


Fig. 3 SEM micrograph of typical PSG from bottom

illustrated in Fig. 5 [8]. In extremely scaled transistors atomic scale channel interface roughness illustrated in Fig. 6 [9] and corresponding oxide thickness and body thickness variations [10] can become important source of statistical variability.

3 Statistical Variability in Advanced CMOS Devices

The simulation results presented in this chapter were obtained using the Glasgow statistical 3D device simulator, which solves the carrier transport equations in the drift-diffusion approximation with Density Gradient (DG) quantum corrections

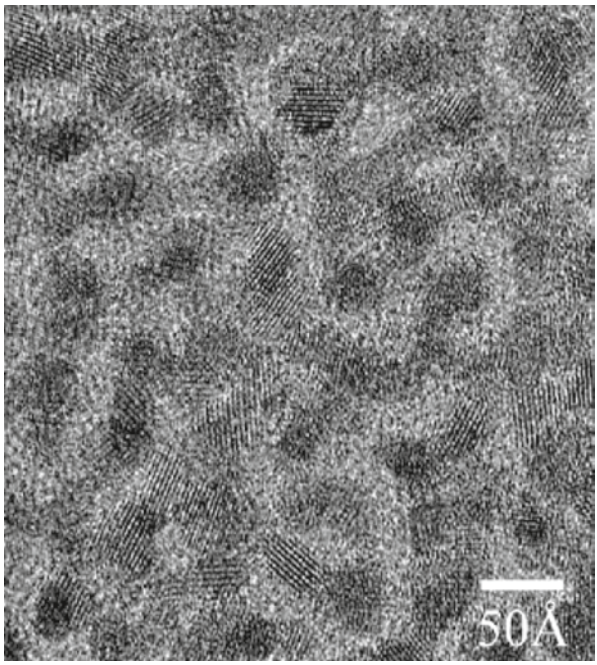


Fig. 4 Granularity in HfON high-k dielectrics. (Sematech)

[11]. In the simulations, the RDD are generated from continuous doping profile by placing dopant atoms on silicon lattice sites within the device S/D and channel regions with a probability determined by the local ratio between dopant and silicon atom concentration. Since the basis of the silicon lattice is 0.543 nm a fine mesh of

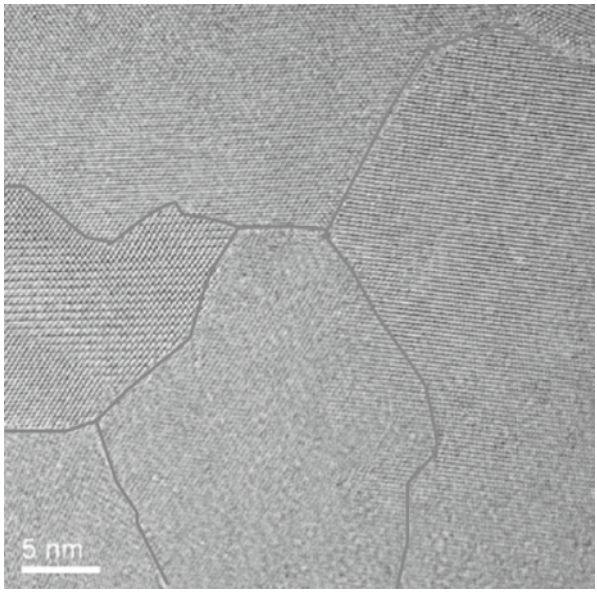


Fig. 5 Metal granularity causing gate work-function variation

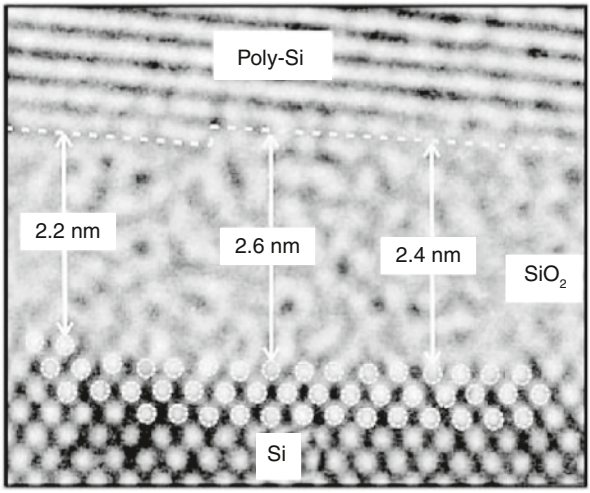


Fig. 6 Interface roughness. (IBM)

0.5 nm is used to ensure a high resolution of dopant atoms. However, without considering quantum mechanical confinement in the potential well, in classical simulation, such fine mesh leads to carrier trapping at the sharply resolved Coulomb potential wells generated by the ionised discrete random dopants. In order to remove this artifact, the DG approach is employed as a quantum correction technology for both electrons and holes [11].

The LER illustrated in Fig. 4 is introduced through 1D Fourier synthesis. Random gate edges are generated from a power spectrum corresponding to a Gaussian autocorrelation function [9], with typical correlation length Λ=30 nm and root-mean-square amplitude Δ=1.3 nm, which is the level that is achieved with current lithography systems. The quoted in the literature values of LER are equal to 3Δ. The procedure used for simulating PGG involves the random generation of poly-grains for the whole gate region [3]: a large atomic force microscope image of polycrystalline silicon grains illustrated at the top of Fig. 3 has been used as a template and the image is scaled according to the average grain diameter (65 nm in the following simulations). Then the simulator imports a random section of the grain template image that corresponds to the gate dimension of the simulated device, and along grain boundaries, the applied gate potential in the polysilicon is modified in a way that the Fermi level remains pinned at a certain position in the silicon bandgap. In the worst case scenario the Fermi level is pinned in the middle of the silicon gap. The impact of polysilicon grain boundary variation on device characteristics is simulated through the pinning of the potential in the polysilicon gate along the grain boundaries. The individual impact of RDD, LER and PSG on the potential distribution in a typical 35 nm bulk MOSFET is illustrated in Figs. 7, 8 and 9 respectively.

The validation of our simulation technology is done in comparison with measured statistical variability data in 45 nm low power CMOS transistors [12]. The simulator was adjusted to match accurately the carefully calibrated TCAD device simulation results of devices without variability by adjusting the effective mass parameters involved in DG formalism, and the mobility model parameters.

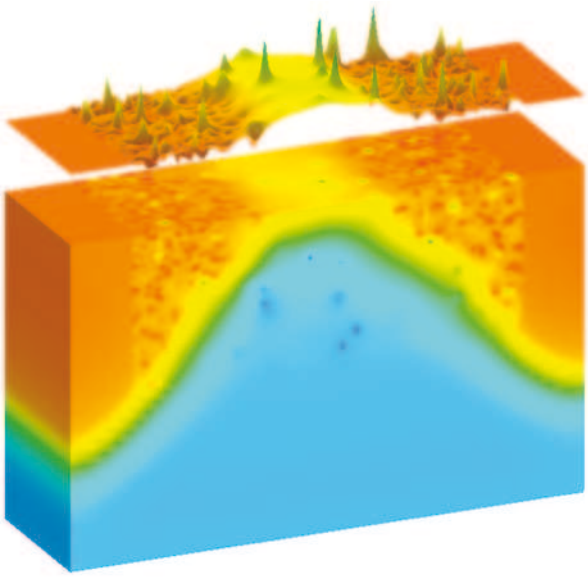


Fig. 7 Potential distribution in a 35 nm MOSFET subject to RDD

The potential distributions in the n- and p-channel transistors in Fig. 7, at gate voltage equal to the threshold voltage and low drain voltage of 50 mV, is illustrated in Fig. 10. In the n-channel transistor RDD, LER and PSG are considered simultaneously while in the p-channel transistor only RDD and LER are considered. The electron concentration at the interface of the two transistors is mapped on top of

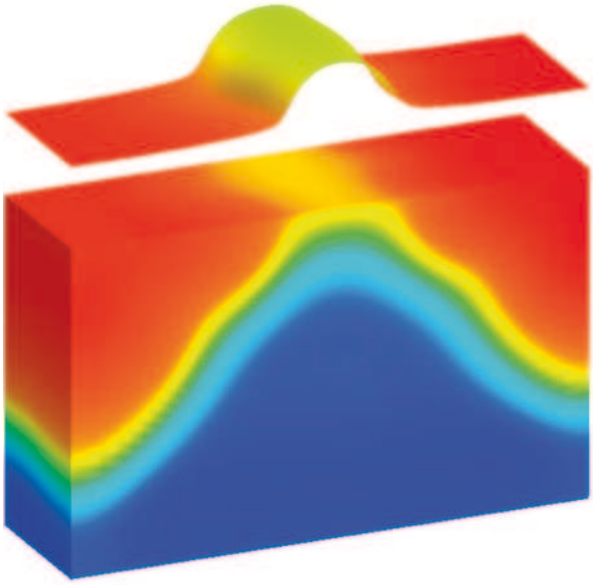


Fig. 8 Potential distribution in a 35 nm MOSFET subject to LER

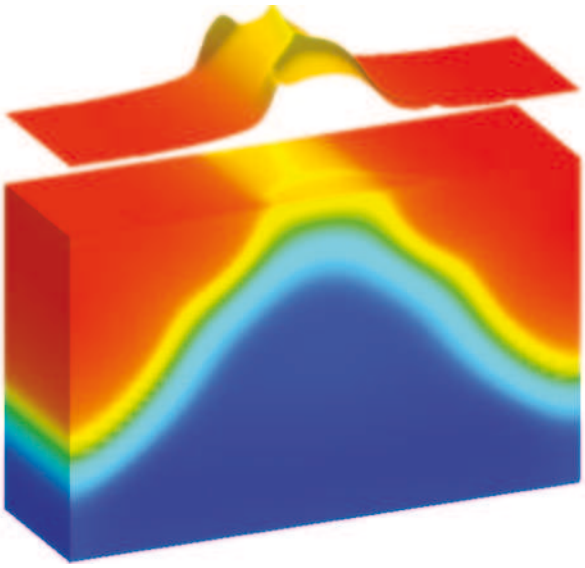


Fig. 9 Potential distribution in a 35 nm MOSFET subject to PSG

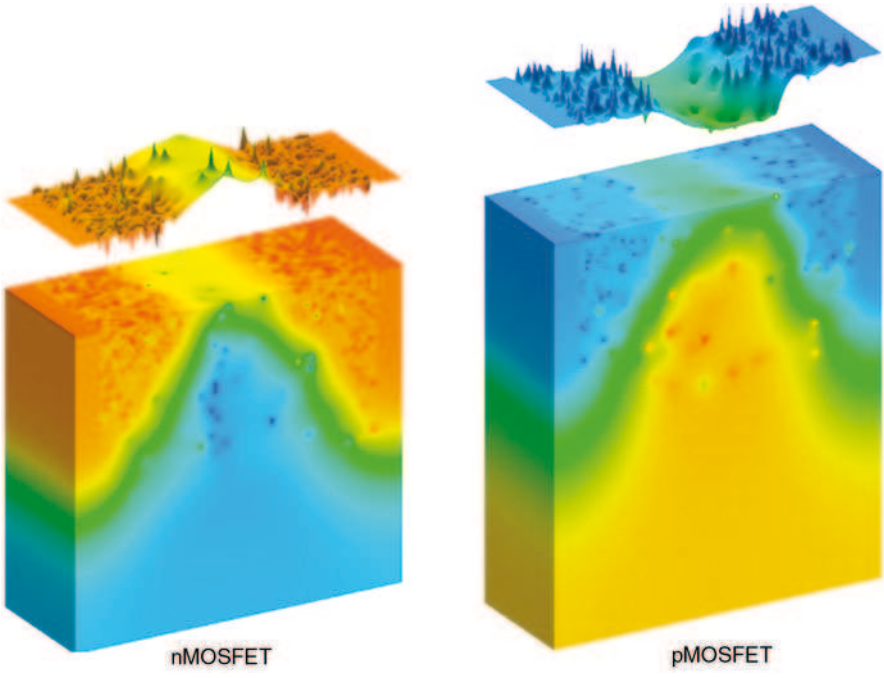


Fig. 10 *Top:* electron (*left*) and hole (*right*) concentration distribution at the interface; *Bottom:* potential distribution

Table 1 σV_T introduced by individual and combined sources of statistical variability

	n-channel MOSFET		*p*-channel MOSFET	
	σV_T (mV) (V_{DS}=0.05 V)	σV_T (mV) (V_{DS}=1.1 V)	σV_T (mV) (V_{DS}=0.05 V)	σV_T (mV) (V_{DS}=1.1 V)
RDD	50	52	51	54
LER	20	33	13	22
PSG	30	26	–	–
Combined	62	69	53	59
Experimental	62	67	54	57

the potential distributions. For example acceptors in the channel of the n-channel transistor create sharp localized potential barrier for the electrons near the interface but act as potential wells for the holes in the substrate. In the same time the donors in the source/drain regions create sharp potential well for electrons.

The simulation results for the standard deviation of the threshold voltage σV_T introduced by individual and combined sources of statistical variability are compared with the measured data in Table 1. In the n-channel MOSFET case the accurate reproduction of the experimental measurements necessitates the assumption that, in addition to RDD and LER, the PSG related variability has to be taken into account. Good agreement has been obtained assuming that the Fermi level at the n-type poly-Si gate grain boundaries is pinned in the upper half of the bandgap at approximately 0.35 eV below the conduction band of silicon.

However, in the p-channel MOSFET case the combined effect of just the RDD and LER is sufficient to reproduce accurately the experimental measurements. The reason for this is the presence of acceptor type interface states in the upper half of the bandgap which pin the Fermi level in the case of n-type poly-Si, and the absence of corresponding donor type interface states in the lower part of the bandgap which leaves the Fermi level unpinned in the case of p-type poly-Si [13].

In order to foresee the expected magnitude of statistical variability in the future we have studied the individual impact of RDD, LER and PSG on MOSFETs with gate lengths 35, 25, 18, 13 and 9 nm physical gate length. We also compare the results with the statistical variability introduced in the same devices when RDD, LER and PSG are introduced in the same devices simultaneously. The scaling of the simulated devices is based on a 35 nm MOSFET published by Toshiba [14] against which our simulations were carefully calibrated. The scaling closely follows the prescriptions of the ITRS in terms of equivalent oxide thickness, junction depth, doping and supply voltage. The intention was also to preserve the main features of the reference 35 nm MOSFET and, in particular, to keep the channel doping concentration at the interface as low as possible. Figure 11 shows the structure of the scaled devices. More details about the scaling approach and the characteristics of the scaled devices may be found in [11].

Figure 12 compares the channel length dependence of σVT introduced by random dopants, line edge roughness and poly-Si grain boundaries with Fermi level

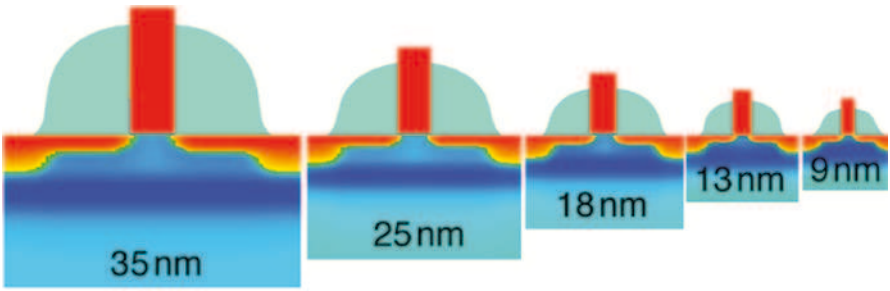


Fig. 11 Examples of realistic conventional MOSFETs scaled from a template 35 nm device according to the ITRS requirements for the 90, 65, 45, 32 and 22 nm technologies, obtained from process simulation with Taurus Process

pinning. The average size of the polysilicon grains was kept at 40 nm for all channel lengths. Two scenarios for the magnitude of LER were considered in the simulations. In the first scenario the LER values decrease with the reduction of the channel length following the prescriptions of the ITRS 2003 of 1.2, 1.0, 0.75, and 0.5 nm for the 35-, 25-, 18-, and 13-nm channel length transistors, respectively. In this case the dominant source of variability at all channel lengths are the random discrete dopants. The variability introduced by the polysilicon granularity is similar to that introduced by random discrete dopants for the 35 nm and 25 nm MOSFETs, but at shorter channel lengths the random dopants take over. The combined effect of the three sources of variability is also shown in the same figure. In the second scenario LER remains constant and equal to its current value of approximately 4 nm ($\Delta = 1.33$ nm). The results for the 35 nm and the 25 nm MOSFETs are very similar to the results with scaled LER but below 25 nm channel length LER rapidly becomes the dominant source of variability.

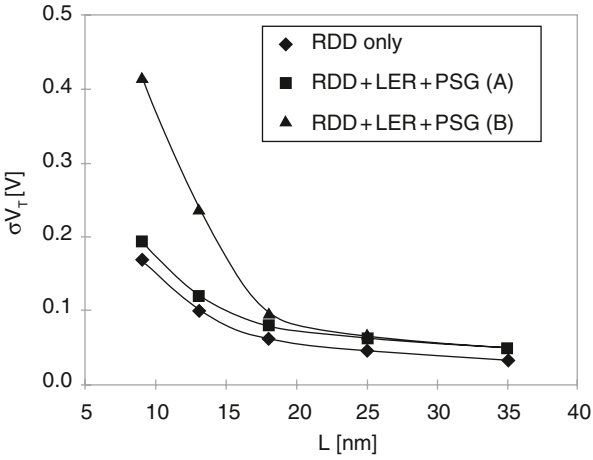


Fig. 12 Channel length dependence of σV_T introduced by random dopants, line edge roughness and poly-Si granularity: *(A)* LER scales according ITRS; *(B)* LER = 4 nm

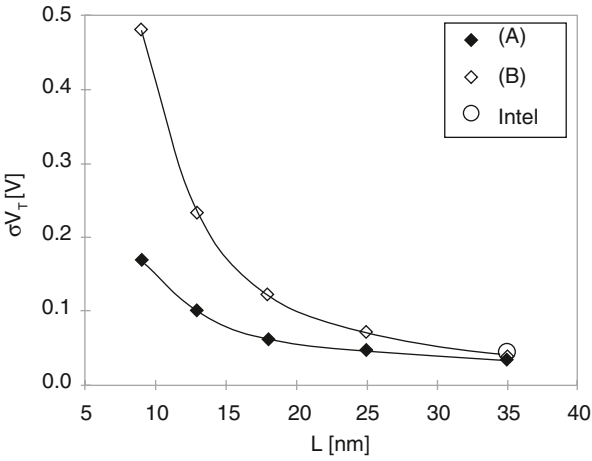


Fig. 13 Channel length dependence of σV_T introduced by random dopants, line edge roughness and poly-Si granularity: *(A)* t_{ox} scales according ITRS; *(B)* $t_{ox} = 1$ nm

Figure 12 is analogous to Fig. 13 exploring the scenario of the oxide thickness, which is difficult to scale further. The LER is scaled according to the ITRS requirements listed above. Even with the introduction of high-k gate stack it is likely to remain stagnated at 1 nm. This will lead to an explosion in the threshold voltage variability for bulk MOSFETs with physical channel length below 25 nm.

4 Statistical Compact Model Strategy

It is very important to be able to capture the simulated or measured statistical variability in statistical compact models since this is the only way to communicate this information to designers [15]. Previous research on statistical compact model identification was focused mainly on variability associated with traditional process variations resulting from poor control of critical dimensions, layer thicknesses and doping clearly related to specific compact model parameters [16].

Unfortunately, the current industrial strength compact models do not have natural parameters designed to incorporate seamlessly the truly statistical variability associated with RDD, LER, PGG and other relevant variability sources. Despite some attempts to identify and extract statistical compact model parameters suitable for capturing statistical variability introduced by discreteness of charge and matter this remains an area of active research [16, 17]. Figure 14 shows the spread in ID–VG characteristics obtained from 'atomistic' simulator due to the combined effect of RDD, LER and PGG.

We use the standard BSIM4 compact model to capture the information for statistical variability obtained from full 3D physical variability simulation. The statistical extraction of compact model parameters is done in two stages [17].

In the first stage, one complete set of BSIM4 parameters is extracted from the I–V characteristics of 'uniform' (continuously doped, no RDD, LER and PGG) set of devices with different channel lengths and widths and process flow identical to the one of the 35 nm testbed transistor (Fig. 15). Target current voltage char-

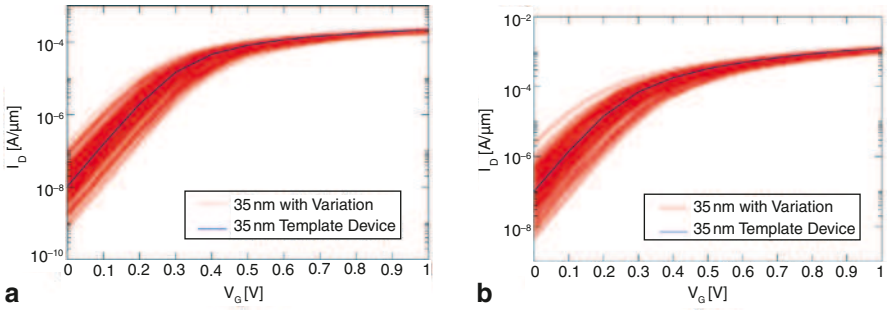


Fig. 14 Variability in the current voltage characteristics of a statistical sample of 200 microscopically different 25 nm square (W = L) n-channel MOSFETs at **a** $V_D = 50$ mV and **b** $V_D = 1$ V

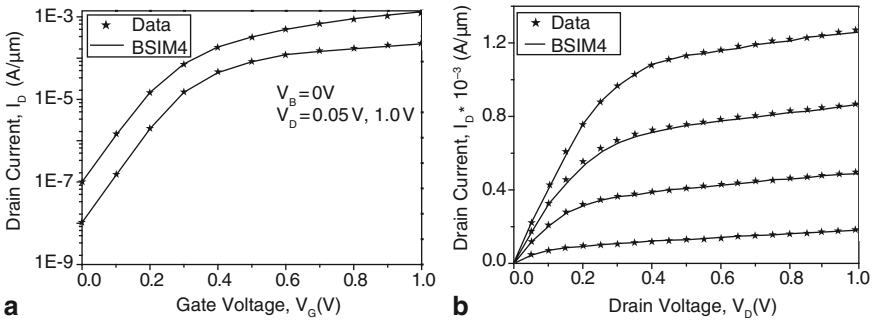


Fig. 15 Comparison of **a** I_D–V_G characteristics, **b** I_D–V_D for characteristics for 'uniform' devices obtained from physical simulations and BSIM

acteristics are simulated over the complete device operating range and parameter extraction strategy combining group extraction and local optimization is employed. Figure 16 compares the corresponding BSIM4 generated I–V characteristics of the 35 nm transistor with the original device characteristics obtained from our 'atomistic' simulator. The RMS error for this extraction is 2.8%.

At the second stage, we re-extracted a small carefully chosen subset of the BSIM4 model parameters from the physically simulated characteristics of each microscopically different device in the statistical ensemble keeping the bulk of the BSIM parameters unchanged. The transfer (ID–VG) characteristics at low and high drain bias are used as extraction target at this stage. The seven re-extracted model parameters are L_{pe0}, R_{dswmin}, N_{factor}, V_{off}, A_1, A_2 and D_{sub}.

Figure 16 compares the BSIM4 results for the worst, the best and the typical devices with the physically simulated device characteristics. The good match between the BSIM results and the physically device characteristics validates the choice of seven parameters which guaranty high accuracy of the compact model over the whole statistical range of physically simulated device characteristics. Figure 17 illustrates the distribution in the error of the 200 statistically different BSIM4 cards depending on the density of the data points in the ID–VG characteristics used as

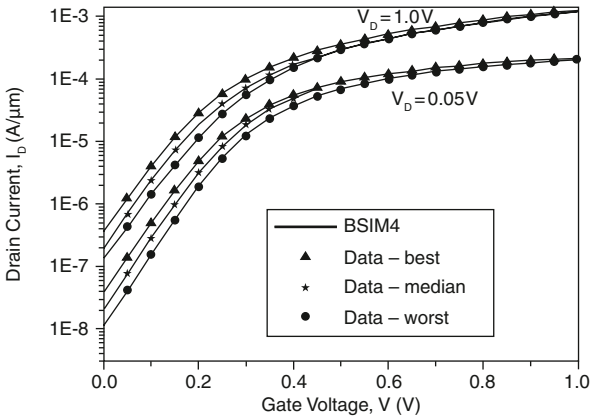


Fig. 16 Comparison of physical and simulation results obtained after second stage statistical extraction for the devices with best, worst and median extraction error

targets in the statistical compact model extraction. Relatively few simulation data points are needed to extract high accuracy statistical compact model parameters.

Above statistical compact model extraction and 3D physical device simulation are based on minimum size square device (W=L), which has the maximum magnitude of statistical variation. However, in real circuits, most of devices have multi-width/length ratio value. In order to reproduce the right statistical behavior of such devices in statistical circuit simulation, wider device are constructed by connecting in parallel minimum size devices randomly selected from the statistic compact model library.

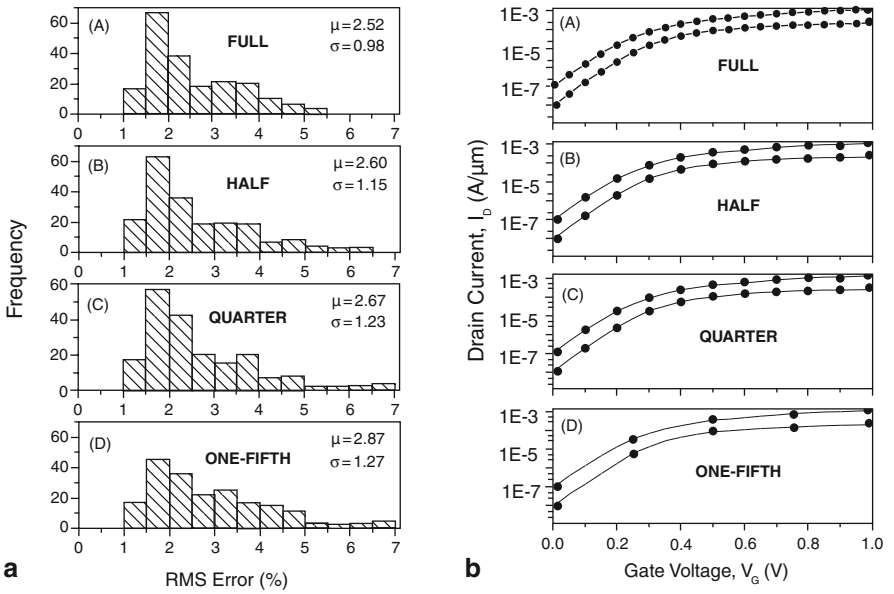


Fig. 17 Distribution of the statistical error as a function of a different number of data points on the target I_D–V_G characteristics used in the statistical compact model extraction

5 Impact of Statistical Variability on SRAM

One area where the digital and the analogue design expertise overlap is in the design of SRAM and particularly SRAM cells which are extremely sensitive to statistical variability. The statistical circuit simulation methodology described in the previous section, which can transfer all the fluctuation information obtained from 3D statistical device simulations into circuit simulation, is employed to investigate the impact of RDF on 6T and 8T SRAM stability for the next three generations of bulk CMOS technology. In the following discussions, we use 25, 18 and 13 nm channel length transistors from Sect. 3.

Currently, 6T SRAM is the dominant SRAM cell architecture in SoC and microprocessors. However, the disturbance of bit lines on the data retention element during read access makes the 6T cell structure vulnerable to statistical variability, which in turn will have a huge impact on 6T SRAM's scalability. The functionality of SRAM is determined by both static noise margin (SNM) defined as the minimum dc voltage necessary to flip the state of the cell and the write noise margin (WNM) defined as the DC noise voltage needed to fail to flip a cell during a write period. The meaning of SNM and WNM is defined in Fig. 18. Figure 19 illustrates the statistical nature of SNM and WNM in the presence of statistical transistor variability.

The magnitude of WNM in SRAM is mainly determined by the load and access transistors illustrated in the 6T SRAM schematics inset in Fig. 20. Since they are the smallest transistors in an SRAM cell, the WNM variation will be larger than the SNM variation. However, the mean value of WNM is much larger than its SNM counterpart. Previous studies [18] suggested that under normal circumstances the limiting factor for the operation of bulk 6T SRAM cells is SNM.

The statistical circuit simulation results for the distribution of SNM for 6T SRAM with the cell ratio of 2 is shown in Fig. 20, while the schematic of a 6T

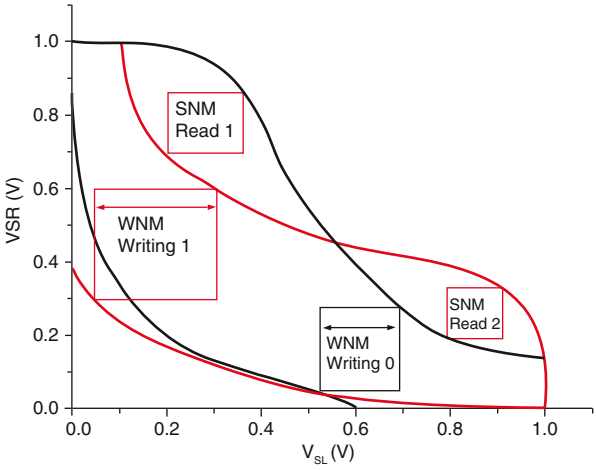


Fig. 18 Static voltage transfer characteristics and definition of SNM and WNM

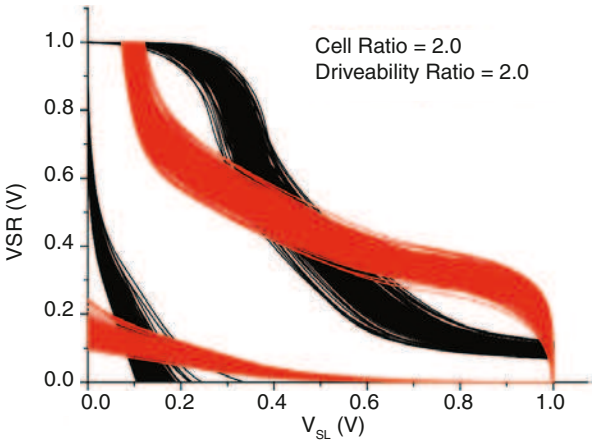


Fig. 19 Statistical behavior of the SNM and WNM of SRAM made of 25 nm bulk MOSFETs subject to RDD

SRAM cell with the bias configuration at the initiation of read operation is illustrated in the inset. The cell ratio is defined as the ratio of the driver transistor to access resistor channel widths. For the 13 nm generation, around 2% of SRAM cells with cell ratio of 2 are not readable even under ideal conditions since their SNM values are negative. The standard deviation σ of SNM, normalized by the

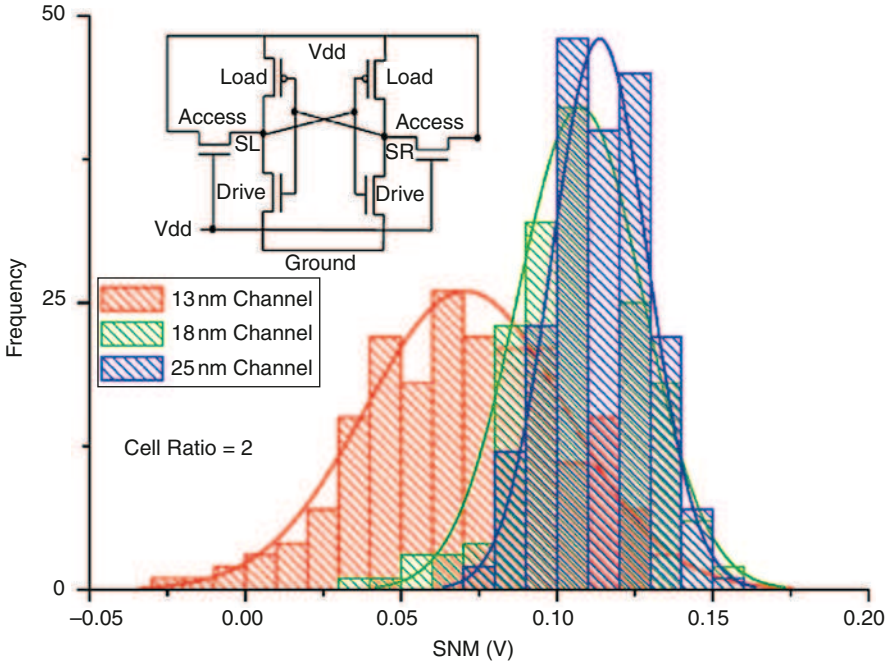


Fig. 20 Distribution of 6T SRAM SNM over an ensemble of 200 SRAM cells for 25, 18 and 13 nm generations

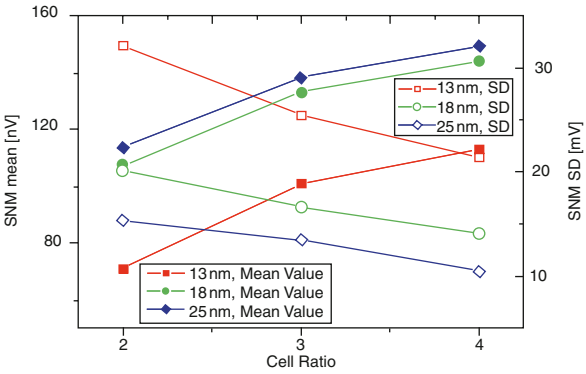


Fig. 21 Mean value and SD of 6T SRAM SNM against SRAM cell ratio

average value μ of the SNM increases from 13% for the 25 nm generation, to 19% for the 18 nm generation, and reaches 45% for the 13 nm generation. As a guideline, $\mu - 6\sigma$ is required to exceed approximately 4% of the supply voltage to achieve 90% yield for 1 Mbit SRAM's. Although the 25 nm generation is three times better with respect to SNM fluctuation performance compared to the 13 nm generation at cell ratio of 2, it still cannot meet the "$\mu - 6\sigma$" test on yield control.

Increasing the cell ratio is the most straightforward way to improve SNM performance of a SRAM cell [19]. The mean value (μ) and the standard deviation (σ) of SNM at different SRAM cell ratios are shown in Fig. 21, which clearly illustrates the benefits of larger cell ratio. For the 25 nm transistors, the increase of the cell ratio from 2 to 4 results in nearly two times NSD of SNM improvement, and the "$\mu - 6\sigma$" criterion is met at a cell ratio of 3. Although the 18 nm and 13 nm generations follow a similar trend, a cell ratio of 4 is required for 18 nm technology in order to achieve reasonable yield. For the 13 nm generation, a cell ratio of 4 can only achieve the level of performance of the 18 nm generation at a cell ratio of 2. At the same time, the increase in the cell ratio tends to degrade the WNM. Therefore, the approach of increasing the cell ratio becomes less attractive for extremely scaled devices even from the prospect of the yield.

6 Conclusions

The statistical variability introduced by discreteness of charge and matter has become one of the major concerns for the semiconductor industry. More and more the strategic technology decisions that the industry will be making in the future will be motivated by the desire to reduce statistical variability. The useful life of bulk MOSFETs, from statistical variability point of view, can be extended below the 20 nm technology mark only if the line edge roughness and the equivalent oxide thickness could be successfully scaled to the required values. SRAM which uses

minimum channel width transistors is the most sensitive part of the integrated systems in respect of statistical variability and needs special care and creative design solution in order to take full advantage from scaling in present and future technology generations.

References

1. G.E. Moore, Progress in Digital Electronics, in *Technical Digest of the Int'l Electron Devices Meeting*. Washington DC, Dec 1975, p. 13
2. K. Bernstein, D.J. Frank, A.E. Gattiker, W. Haensch, B.L. Ji, S.R. Nassif, E.J. Nowak, D.J. Pearson, N.J. Rohrer, IBM J. Res. Dev. **50**, 433 (2006)
3. A.R. Brown, G. Roy, A. Asenov, Poly-Si gate related variability in decananometre MOSFETs with conventional architecture. IEEE Trans. Electron. Dev. **54**, 3056 (2007)
4. B.-J. Cheng, S. Roy, A. Asenov, The impact of random dopant effects on SRAM cells, in *Proc. 30th European Solid-State Circuits Conference (ESSCIRC)*, Leuven, 2004, p. 219
5. A. Agarwal, K. Chopra, V. Zolotov, D. Blaauw, Circuit optimization using statistical static timing analysis, in *Proc. 42nd Design Automation Conference*, Anaheim, 2005, p. 321
6. A. Asenov, Random dopant induced threshold voltage lowering and fluctuations in sub 0.1 micron MOSFETs: A 3D 'atomistic' simulation study. IEEE Trans. Electron. Dev. **45**, 2505 (1998)
7. A. Asenov, S. Kaya, A.R. Brown, Intrinsic parameter fluctuations in decananometre MOSFETs introduced by gate line edge roughness. IEEE Trans. Electron. Dev. **50**, 1254 (2003)
8. J.R. Watling, A.R. Brown, G. Ferrari, J.R. Babiker, G. Bersuker, P. Zeitzoff, A. Asenov, Impact of High-k on transport and variability in nano-CMOS devices. J. Comput. Theo. Nanosci. **5**(6), 1072 (2008)
9. A. Asenov, S. Kaya, J.H. Davies, Intrinsic threshold voltage fluctuations in decanano MOSFETs due to local oxide thickness variations. IEEE Trans. Electron. Dev. **49**, 112 (2002)
10. A.R. Brown, J.R. Watling, A. Asenov, A 3-D atomistic study of archetypal double gate MOSFET structures. J. Comput. Electron. **1**, 165 (2002)
11. G. Roy, A.R. Brown, F. Adamu-Lema, S. Roy, A. Asenov, Simulation study of individual and combined sources of intrinsic parameter fluctuations in Conventional Nano-MOSFETs. IEEE Trans. Electron. Dev. **52**, 3063–3070 (2006)
12. A. Cathignol, B. Cheng, D. Chanemougame, A.R. Brown, K. Rochereau, G. Ghibaudo, A. Asenov, Quantitative evaluation of statistical variability sources in a 45 nm technological node LP N-MOSFET. IEEE Electron. Dev. Lett. **29**, 609 (2008)
13. A. Asenov, A. Cathignol, B. Cheng, K.P. McKenna, A.R. Brown, A.L. Shluger, D. Chanemougame, K. Rochereau, G. Ghibaudo, Origin of the asymmetry in the magnitude of the statistical variability of n- and p-Channel poly-Si gate bulk MOSFETs. IEEE Electron. Dev. Lett. **29**, 913–915 (2008)
14. S. Inaba, K. Okano, S. Matsuda, M. Fujiwara, A. Hokazono, K. Adachi, K. Ohuchi, H. Suto, H. Fukui, T. Shimizu, S. Mori, H. Oguma, A. Murakoshi, T. Itani, T. Iinuma, T. Kudo, H. Shibata, S. Taniguchi, M. Takayanagi, A. Azuma, H. Oyamatsu, K. Suguro, Y. Katsumata, Y. Toyoshima, H. Ishiuchi, High performance 35 nm gate length CMOS with NO oxynitride gate dielectric and Ni salicide. IEEE Trans. Electron. Dev. **49**, 2263 (2002)
15. M. Denais, V. Huard, C. Parthasarathy, G. Ribes, F. Perrier, N. Revil, A. Bravaix, Interface trap generation and hole trapping under NBTI and PBTI in advanced CMOS technology with a 2-nm gate oxide. IEEE Trans. Dev. Mater. Reliab. **4**, 715 (2004)
16. C.C. McAndrew, Efficient statistical modeling for circuit simulation, in *Design of Systems on a Chip: Devices and Components*, ed. by R. Reis, J. Jess (Kluwer, Dordrecht, 2004), p. 97. ISBN 978-1-4020-7929-0

17. K. Takeuchi, M. Hane, Statistical compact model parameter extraction by direct fitting to. IEEE Trans. Electron. Dev. **55**, 1487 (2008)
18. B. Cheng, S. Roy, G. Roy, F. Adamu-Lema, A. Asenov, Impact of intrinsic parameter fluctuations in decanano MOSFETs on yield and functionality of SRAM cells. Solid-State Electron. **49**, 740 (2005)
19. B. Cheng, S. Roy, A. Asenov, CMOS 6-T SRAM cell design subject to 'atomistic' fluctuations. Solid-State Electron. **51**, 565 (2007)

Advanced Physical Design in Nanoscale Analog CMOS

Lanny L. Lewyn

1 Introduction

At the 45 nm technology node, the problem of realizing manufacturable high-performance analog circuits is becoming increasingly more severe. In the pre-nanoscale era, successful LVS-clean post-layout circuit simulation and DRC-clean physical design were adequate design verification steps. Presently, lithographic pattern fidelity, mechanical and electrical stress, process variability, increased wiring parasitics, and many other well documented effects [1] have combined to decrease the probability that a 45 nm circuit will still perform with adequate margins. Unfortunately, this is true even after all post-layout design verification steps have been completed. Pre-tapeout physical-design verification steps that used to take 2 weeks to complete can now take 2 months, or more.

While many of the physical effects which have a strong influence on analog circuit performance can be modeled using technology-aware (TCAD) circuit simulators, these simulators are seldom used by analog CMOS circuit designers. The drive for the semiconductor industry to keep pace with Moore's law, fueled by huge investments in process development and capital equipment, has resulted in process technology outpacing circuit and physical design verification tools. At the beginning of the 100 nm technology era this difference in pace was called 'the reliability gap.' This gap was identified as discrepancies between circuit reliability and DRC-clean physical design arose [2]. The discrepancy of primary concern was short electromigration-limited (EM) lifetimes.

EM verification tools are now in common use, and are quite accurate. However, they are only useful in the post-layout extraction phase of the design. With each advance in technology node, EM coefficients decrease and line widths decrease. The results is an inverse second-order dependence on wire current limits with technology-node minimum feature size (F) [3]. This trend requires that different strategies for both device and interconnect physical design be employed at 32 nm and beyond.

L. L. Lewyn (✉)
Lewyn Consulting Inc., Laguna Beach, CA, USA
e-mail: lanny@pacbell.net

H. Casier et al. (eds.), *Analog Circuit Design,*
DOI 10.1007/978-94-007-0391-9_3, © Springer Science+Business Media B.V. 2011

With the continued use of 193 nm DUV down to the 22 nm technology node, the problem of lithographic pattern fidelity (affecting gate length *L*, gate width *W*, etc.) has become more severe and also requires changes in analog physical design strategy. To complicate matters, pattern fidelity effects are highly dependent on distance to nearby structures or devices. Stress engineering required to increase mobility in deep nanoscale devices has not only increased variability in analog device-pair matching, but also affected the performance of single logic inverters where distance to nearby device oxide-definition patterns is too great, or has large variations.

As the EDA industry strives to keep pace with advancements in process technology, the CMOS IC verification tools are identifying a wider range of circuit and physical design deficiencies. While the EDA tools have shown considerable progress, problems found in post-layout verification can result in significant additional redesign and re-layout time. Circuit and layout redesign effort wastes the semiconductor industries most valuable commodity, time to market.

Because of the severe impact of electromigration effects, pattern fidelity limitations, process variability, and interconnect parasitics on circuit performance (DFP), reliability (DFR), and manufacturability (DFM) at the 32 nm node and beyond, major IC design firms are requiring significantly more participation from circuit design engineers in the physical design. For example, analog circuits designed with wide variations in designated device widths (W) exhibit pattern irregularity, force varied interconnect metal layer combinations, and impact device performance as a result of local stress-gradient irregularity.

Unfortunately, the participation of circuit design engineers in the physical design is becoming more difficult as the number of basic design rules has more than doubled. The number of optional, or recommended design rules has gone up by an order of magnitude in comparison to pre-nanoscale technologies. An objective of this paper is to identify many factors adversely affecting successful nanoscale analog physical design. However the primary objective is to propose methods for mitigating or overcoming physical design problems so that the future of deep-nanoscale analog CMOS looks a bit brighter. One key to the basic approach is to propose certain restrictions on device design at the circuit design stage and to achieve increased regularity in physical design at layout.

2 Pre-Layout Simulation

In pre-layout simulation at the 45 nm node and beyond, device effective W, L, Vt, and *gm* in simulation can be modified by including several geometry-dependent factors. Prior to the nanoscale era of technology, entry of device junction and some interconnect parasitics was sufficient. At the 100 nm technology node, the distance to well edge (WPE effect [4]) was required to modify device Vt. Average distance from gate to shallow trench isolation edge (STI effect [5]) modified both Vt and mobility to account for stress effects. It was also common practice to add global, or specific local interconnect parasitic capacitances and resistance.

At 45 nm and beyond, additional stress and patterning effects have been added to modify device *W* and *L*. These include distance between oxide definition areas and distance from gate to other gates (more than one). Yet, many other geometry-dependant factors important to small-signal analog design and I/O driver design, such as distance to overlying metal, are not yet modeled (even in post-layout simulation). Another complication arises as the number of geometry-dependant simulation parameters increases. If the default values for these parameters are not sufficient to achieve good simulation results, then accurate values must be entered manually, or by developing automated methods.

The practice of employing a number of different device sizes, such as large variations in gate widths, makes the value of oxide-definition separation distances (in the *W* direction) highly layout dependant and difficult to estimate in the pre-layout phase of circuit design. Restricting the possible values of *W* and relying more on varying device *M* is a common practice in bipolar technology and helps increase device pattern regularity in analog CMOS design. Restricting the number of gate fingers (F) to 2 in a single device can bring values for STI parameters SA, SB very close to default pre-layout simulation values. Restricting device design patterns to achieve oxide definition and poly-spacing regularity as well as achieving regularity and symmetry using other circuit physical design techniques (to mitigate the WPE effect, for example) will be addressed in paragraph 6.

3 Post-Layout Simulation

One of the most important post-layout verification checks in the pre-nanoscale era was for EM. EM was recognized as one of the most important factors affecting IC reliability in the nanoscale era of technology. EM-limited minimum line widths and decreasing via-current capabilities have imposed severe restrictions on nanoscale analog device geometry.

It is no longer practical for the circuit designer to be unaware of EM restrictions on device design at the 45 nm technology node and beyond. If the physical designer has to break a device up into shorter widths and more multiples of devices (M), or fingers (F), to achieve a practical wiring configuration, then pre-layout simulation success is no longer a good prediction of manufacturability.

As post-layout verification tools improve, the ability to predict properties such as poly-corner rounding (*L* modifications) and oxide-definition corner rounding (*W* modifications) improves simulation accuracy and results in improved circuit manufacturability. Unfortunately, for reasons previously stated, finding that physical effects have substantially reduced circuit performance at the later verification stages of the design cycle is something to be avoided. We are entering an era when the circuit designer must be increasingly aware of the physical effects that are consequences of circuit design choices. The alternative is to build in schedule slips for redesign at the later verification phases of the design cycle. As described previously with regard to improving pre-layout simulation accuracy, imposing restrictions on

device design and achieving regularity in the physical design are key methods for avoiding unpleasant surprises at the post-layout verification phase of the design.

Unfortunately, many adverse optical proximity effects which result in poor pattern fidelity still cannot be properly modeled by non-TCAD post-layout verification tools. Simple techniques for mitigating lithographic fidelity problems such as significantly increasing gate extension for analog devices and achieving pitch regularity in device layout will be addressed in paragraph 6. It should be noted that since the choice of mask exposure wavelength has remained at 193 nm (DUV) down through the 22 nm node, gate corner rounding and gate-to-oxide-definition-layer misalignment effects have had a constantly increasing impact on systematic matching errors in nanoscale analog CMOS.

4 Process Variability

In the pre-nanoscale era, total parameter variation was used for slow, typical, and fast-corner circuit simulation models. Total parameter variation was primarily a result of global across-chip, wafer-to-wafer, or lot-to-lot parameter variation. Local (within a circuit) parameter variation was a small fraction of global variation. At highly scaled process nodes, local parameter variation is becoming a substantial, or even greater, fraction of global variation. Many technology factors contribute to this increase in local variation. These factors include device stress, implant range depth, exposure variation (exposure latitude), focus variation (depth of focus), mask feature size variation, and mask misalignment. To make matters worse for the analog circuit designer, the use of reticle resolution enhancement techniques (RET) to correct for optical proximity effects (OPC) result in the mask geometry no longer conforming to the physical layout as shown in Fig. 1 [6]. After the application of OPC, the gate layer patterning (poly at 45 nm) achieves higher CD accuracy in the final silicon as shown in the right panel of Fig. 2.

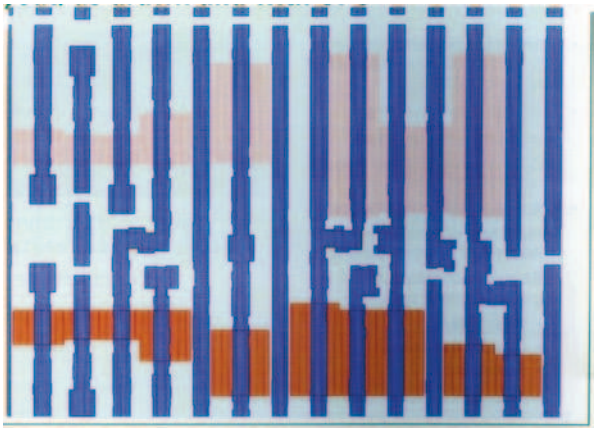


Fig. 1 After OPC, the reticle patterning no longer has a 1:1 correspondence with the layout [6]

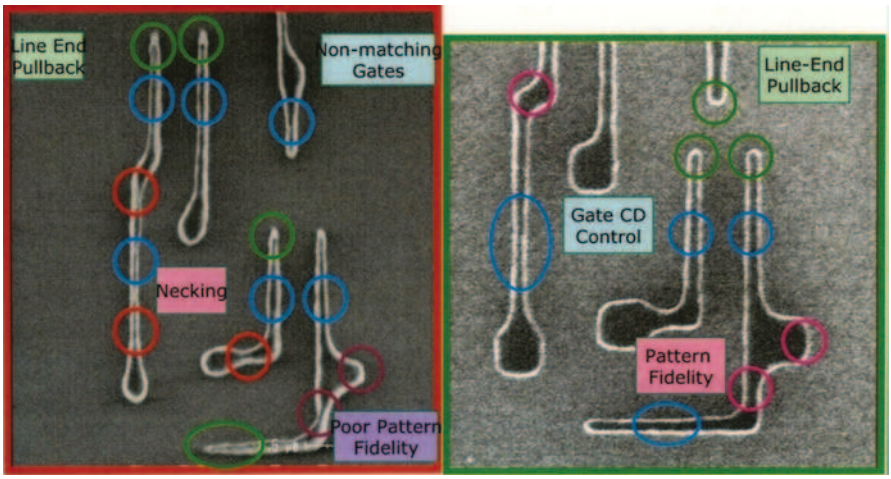


Fig. 2 Following the application of optical proximity correction (OPC) to the reticle, poly layer patterning achieves higher critical-dimension (CD) accuracy in the final silicon (*right panel*) [6]

While lithography effects are responsible for increasing one component of local process variability, stress and reactive-ion-etch (RIE) loading effects have for a long time been significant factors in degrading precision CMOS DAC (C-DAC) accuracy. These effects and methods for mitigating them are addressed in paragraph 7 of this paper.

5 Mitigating Local Mismatch Errors

Local mismatch errors are those which occur at small distances, such as between adjacent devices. These errors include quantum edge or thickness effects and process variations over a small distance. Physical design errors also induce systematic mismatch, such as where distances from gate to substrate tie are asymmetric in the case of high frequency (HF) analog devices.

5.1 Quantum-Effect Mismatch Errors

The effects of quantum variations in oxide thickness, edge roughness, and depletion region carrier concentrations have been cited by several authors as significant sources of mismatch in highly scaled devices. Minimum area (minimum W and L) devices suffer from these mismatch effects in nanoscale CMOS. However, minimum area devices are not used by analog designers to achieve good matching accuracy. In cases where a relatively low (>10 mV) offset is required and device area is large, quantum mismatch effects are not a significant source of error.

5.2 Stress-Induced Mismatch Factors

Threshold voltage matching coefficient *A(Vt)* probably will not get much better than the 2.5 mV/μm industry average value for low-voltage threshold PMOS at 65 nm. It should be noted that: (1) Low threshold devices have a lower bulk implant dose, resulting in perhaps 10% better matching accuracy than standard-threshold devices; (2) the added implants in high threshold devices result in less Vt match accuracy than standard-threshold devices; (3) NMOS may have less Vt match accuracy than PMOS; and (4) at high values of gate drive (Vgs-Vt), the saturation current Idsat matching constant *A(Idsat)* is dominant over *A(Vt)*.

Other process variability factors are beginning to turn *A* values upward at the 45 nm node and beyond. These factors include random components of process variables such as implant range (depth) and stress induced by irregular overlying metal and nearby device patterns within the local (<2 μm) stress field. Recent stress factors include STI and features added to overcome reduced carrier mobility from higher channel doping. These include cap layers over PMOS gates and Ge implants into raised S-D structures, as shown in Fig. 3 [1]. Antenna-effect may cause significantly more mismatch error than sources previously described. However, this source of error is easily overcome by the addition of back-biased diodes to gates with long or high-level metal wiring.

5.3 High-Frequency Device Matching

Nanoscale high-frequency (HF) amplifiers are presently capable of achieving bandwidths in excess of 10 GHz. An additional factor relevant to the matching of HF amplifier or comparator devices is that the distances within each device to the nearby substrate ties must be also matched. Increasing the distance to substrate ties increases local tub (bulk) resistance components R1 and R2 (see Fig. 3). These resistances isolate the device MOS model substrate node from the device subcircuit-symbol Vb node and degrade HF performance.

5.4 Strategy for Matching as Devices are Downscaled

Downscaling by the process general shrink factor maintains constant device *W/L* ratios but generally decreases device area by half for each process generation. Since Vt mismatch is an inverse function of area square-root, it is apparent that downscaling by 2 process generations will increase mismatch by a factor of 2, even if there is no change in the *A* factor. Downscaling while maintaining constant area tends to keep matching from degrading rapidly, but gain-bandwidth (GBW) reductions makes this method practical only for low-frequency current mirror or current-bias

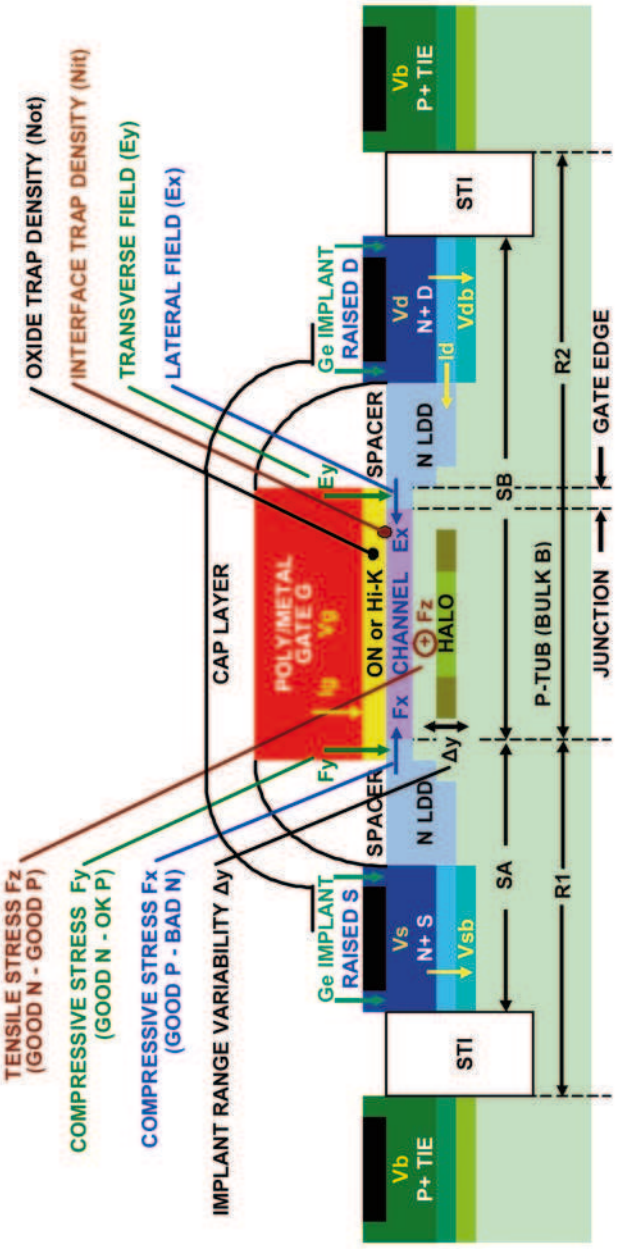


Fig. 3 Nanoscale MOS cross section. NMOS doping polarities are shown, however a typical PMOS cap layer is also shown. In addition to tensile stress from cap layers and *Ge* raised source-drain (*S-D*) implants, dimensions such as distance from source-channel boundary to nearby *STI* (*SA* and *SB*), distance to nearby device oxide-definition layers, and overlying metal patterns within the local (< 2 μm) stress field induce transverse (Fy) and lateral (Fx and Fz) stress

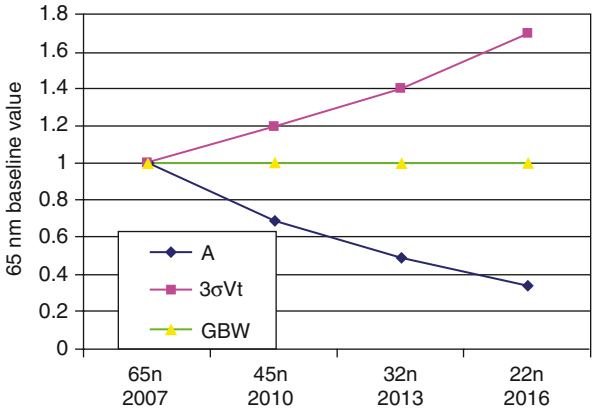


Fig. 4 Constant *GBW* downscaling. Device area *A* is scaled down linearly with *F*. *W/L* is constant as both are scaled in proportion to $F^{-0.5}$, so device current remains constant. There is a moderate increase in *Vt* mismatch

devices. Downscaling devices in circuits where both GBW and matching is important can be done with some loss in performance by scaling *W* and *L* features as $F^{1/2}$, rather than *F* (see Fig. 4.)

6 HF Analog Device Physical Design

Many of the factors contributing to local mismatch errors can be mitigated by adopting a uniform approach to NMOS and PMOS physical design. As previously stated, restricting the range and number of values of device *W* helps increase device pattern regularity in analog CMOS design.

6.1 Restricting the Number of Gate Fingers (F)

Restricting the number of gate fingers (*F*) to 2 in a single device has multiple advantages. In addition to bringing values for STI parameters SA, SB very close to default pre-layout simulation values, it insures a short distance from the gate to the substrate tie. This decreases local tub (bulk) resistance RSUB components R1 and R2, shown in Fig. 3.

The degree to which bulk resistance components is reduced with 2, 4, and 6 finger configurations in a typical 100 nm HF CMOS device is shown in Table 1. There is almost a factor of 2 difference in RSUB between the 2 and 6 figure options. This difference is not important for DC bias devices. However, when attempting to achieve the highest *GBW* possible, even modest decreases in RSUB are useful in improving HF bandwidth.

Table 1 TCAD simulation results from 100 nm device showing variation of RSUB (R1 and R2 average) as number of device fingers are varied. Note that the difference in RSUB approaches a factor 2 between 2 and 6 finger options. This difference is important in the physical design of HF CMOS devices

Basic W = 0.8 μM device structure	Properties	Device width trench to trench (μM)	RSUB to center drain (Ohm)
2 finger NMOS (M2)	Min. tie width	1.3	1250
4 finger NMOS (M4)	Min. tie width	2.3	1667
4 finger NMOS (M4)	Double tie width	2.3	1613
6 finger NMOS (M6)	Double tie width	3.3	2083
2 finger NMOS (M2)	Exact structure	1.3	1316

6.2 Device Optimum Current Density Ids/(W/L) Under Scaling

Several concepts related to circuit design with MOS devices are simplified by understanding the optimum relationship between drain current *Ids* and the number of squares (*W/L*) required in the physical design of the device. This relationship may be expressed simply as device current density *Ids*/(*W/L*) in units of uA/sq. where as-drawn values of *W* and *L* are used to simplify the calculations.

Over many process generations from pre-nanoscale to deep-nanoscale, the current density for NMOS devices has been 15 uA/sq. where high GBW is required and high values (>300 mV) of Vdsat are acceptable. For moderate GBW and Vdsat, the current density is nominally 10 uA/sq. For very low Vdsat (<200 mV) and low GBW, such as in the case of current bias mirrors, 5 uA/sq. is acceptable. The corresponding values for PMOS devices are 5, 3, and 2 uA/sq. respectively. Note that low power (LP) processes have substantially poorer device Vdsat and GBW performance than high performance (HP) processes.

In order to reduce device capacitance, shared-drain physical design approaches are commonly preferred by analog designers in HF device design. As EM physical design constraints force the use of more metal layers for device connections, the adverse effects of multi-layer-metal drain-source capacitance for arrangements such as S-D-S-D-S-D-S should be apparent. In contrast, the arrangement S-D-S, S-D-S has the same number of source electrodes, but fewer drain electrodes. The S-D-S device construction is shown in Fig. 5.

6.3 EM Limitations Forcing Use of Multiple Metal Layers

The main reason that EM limitations are forcing the use of multiple metal layers at the device drain for typical CMOS HF amplifier applications is the increase in current density resulting from lower-capacitance shared-drain device design. Shared-drain devices have twice the current density. For a shared-drain HF NMOS device, the typical difference is 20 uA/sq. vs. 10 uA/sq. for an unshared-drain configuration. TCAD simulations reveal that the incremental increase in parasitic drain-gate

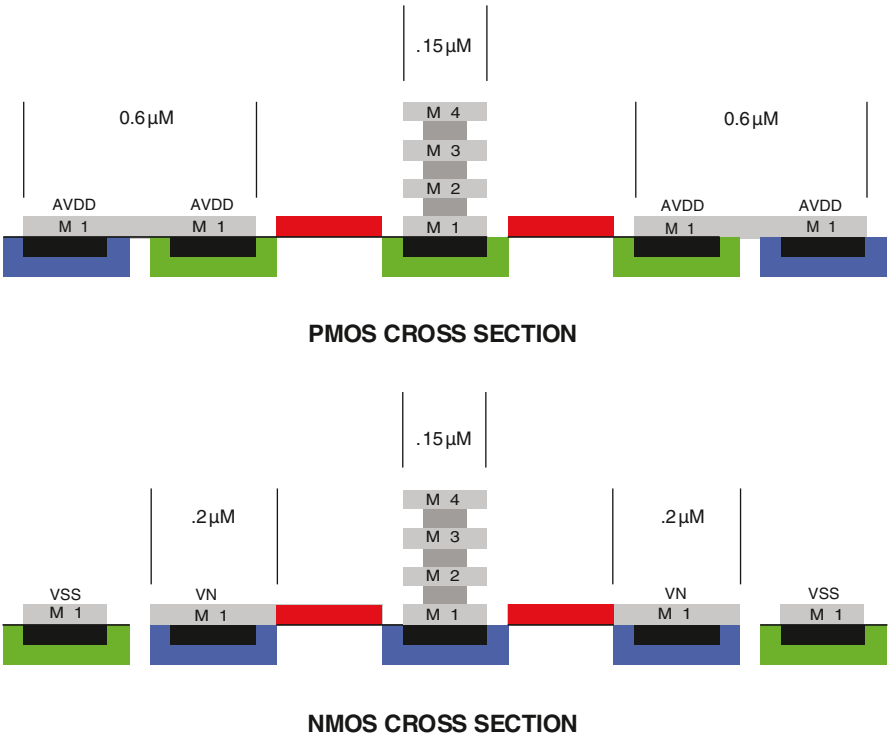


Fig. 5 A cross section example of HF device construction using the S-D-S methodology. Note the proximity and symmetry of the substrate tie locations

capacitance *Cdg* falls off sharply as the number of drain metal layers is increased, so the extra drain metal layers do not pose a serious GBW problem.

At the 22 nm technology node, drain metal patterns must be increased beyond minimum width values to accommodate 20 uA/sq. drain currents as seen in Fig. 6. The wider metal forces additional gate-to-gate spacing and higher drain area. The increased drain junction capacitance is not significant and the increased gate-to-gate spacing allows for the placement of more VIAs. VIA current-density limits are driving HF amplifier drain shapes more strongly than metal EM limits at the 32 nm node.

6.4 Achieving Good Pattern Fidelity in Device Design

To achieve good pattern fidelity in device design, factors such as gate spacing regularity, restrictive gate lengths and finger count, directional uniformity, oxide definition regularity and metal routing placement must be given careful consideration. The device design shown in Fig. 7 is proposed to satisfy many of the important

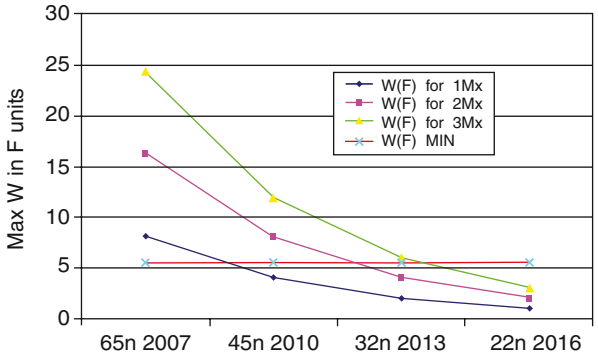


Fig. 6 Maximum device *W* (in F units) for 1, 2 and 3 layers of Mx. Calculations assume an industry average EM current limit for short-length wires of 1 mA/μm at the baseline (65 nm) technology node. When maximum device *W* falls below minimum width required for 2 drain contacts at $W \approx 5.5$ F (*W MIN*—horizontal line), an additional metal layer is required [3]

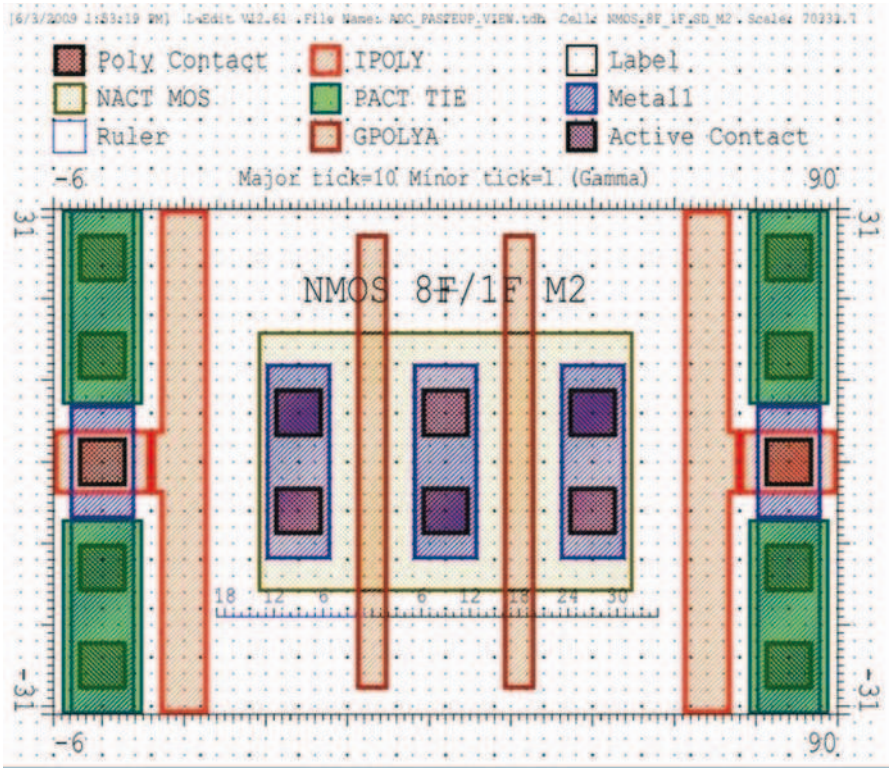


Fig. 7 Basic device design accommodates up to L=2 F, has close tie spacing to achieve good HF performance. Dummy gate connects to metal Vss between vertical tie split to achieve gate pattern uniformity. Dimensionless GAMMA (Γ) units are used where 1Γ=0.01 μm at the 32 nm node. Minimum gate extension is 12 Γ at 45 nm and beyond to avoid gate-end rounding effects

considerations in nanoscale CMOS HF amplifier device design. In the device of Fig. 7, STI spacing values resulting from SA-SB averages are either very close to foundry default model values, or sufficiently regular to be entered into simulation model decks as global parameters.

In deep nanoscale technology, SA-SB average spacing for minimum-length devices may be tuned to achieve optimal *gm* values near the model default values. In that case, it is a performance advantage for devices to be constructed with 2 finger gates. Gate-to-substrate capacitance is not substantially increased by imposing the 2-finger gate restriction. Device gate capacitance will dominate gate M1 wiring *X*-direction interconnect parasitic capacitance by more than an order of magnitude. Therefore the additional *X*-direction array width does not significantly reduce GBW.

Some sacrifice in device area efficiency will result from imposing the maximum F = 2 finger restriction, however this restriction enables extreme regularity in oxide-definition-layer pattern spacing. This regularity is useful in setting up global pre-layout simulation values for *X* and *Y* oxide-definition layer spacing as shown in Fig. 8. In the figure, upper metal layers and PMOS well patterns have been deleted for clarity. However, the WPE effect is virtually eliminated by providing uniformity in PMOS device widths and extending well patterns horizontally into adjacent cells.

Figure 8 shows that in a practical HF amplifier design the use of uniform device dimensioning results in regular patterns for the trench isolation as the device pat-

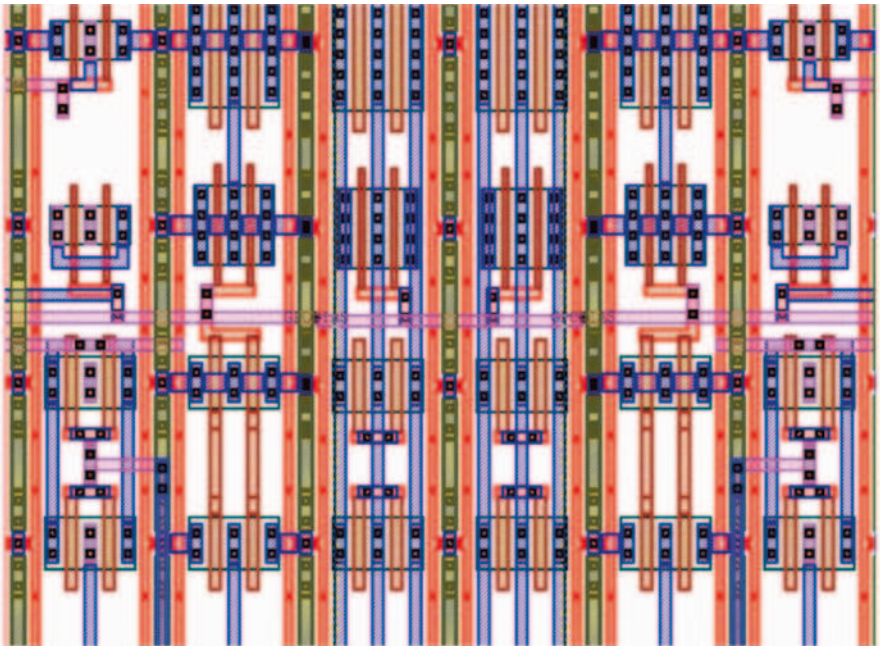


Fig. 8 Uniform 2F MOS device cells are combined in an amplifier cell to provide regular oxide-definition spacing. The spacing scheme allows regular. Low-Cp Mx tracks over STI isolation for signal interconnect and EM-robust, direct interconnect, from NMOS to PMOS drains

terns shown in Fig. 7 overlap. This pattern regularity would not be easily achieved in the case where some devices were constructed with 2-finger gates and others were constructed with 4 or 6 fingers. It is clear that the device pattern uniformity also improves the regularity of the metal routing. Other advantages of device pitch uniformity that contribute to accuracy of precision circuits, such as C-DAC weighting networks, will be discussed in paragraph 7.

7 Mitigating Long-Distance Mismatch Errors in Precision Circuits

Process gradients and variations induced by non-uniform lithographic patterns at distances of 50 μM adversely affect the accuracy of precise capacitor arrays [7]. First-order (linear) process gradients affect ADC poly-poly and MIM capacitor dielectric thickness. Second-order (quadratic) lithographic errors are primarily a result of reactive-ion etch (RIE) loading and affect poly-poly, MIM and fringe-capacitor edge pattern uniformity. Nanoscale fabrication technology has significantly improved precise capacitor edge-definition accuracy, but not significantly reduced the critical 50 μM RIE-loading-effect distance.

Although it is usual practice to place 1 or 2 rows of dummy capacitors around a 2-dimensional precise capacitor array, dummy capacitors placed out to a distance up to several 10 s of microns is not area efficient. In a 16b uncalibrated ADC used in a space-imaging application, process gradient errors are mitigated by amplifier, capacitor, and switch array uniformity in 1 dimension, *X* [8]. Figure 9 shows a portion

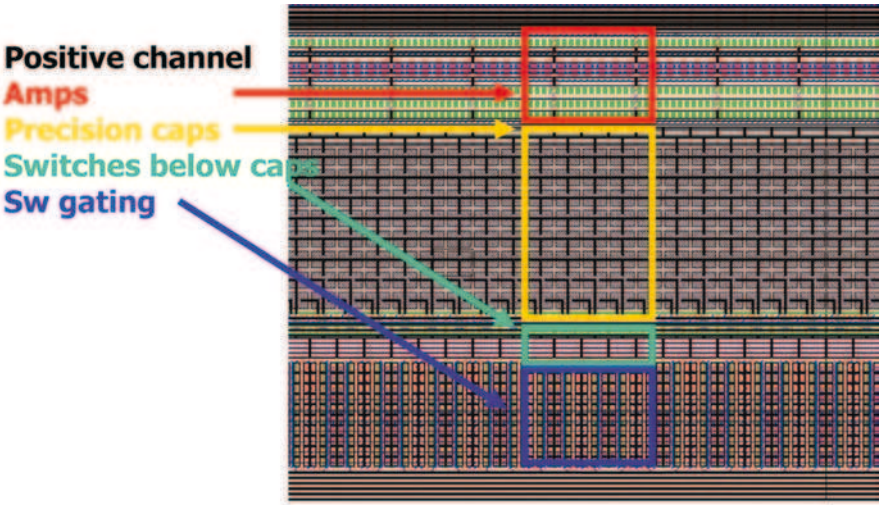


Fig. 9 A portion (approx. 20% of the total *X*-dimension) of an uncalibrated 16 b ADC. The ADC was used to replace the ADC in the Hubble Advanced Camera for Surveys (ACS) on Hubble servicing mission SM-4 in May 2009 [8]

of the ADC MSB poly-poly capacitor-DAC (C-DAC) with unit capacitor selection in only 1 dimension (*X*). Rather than adding many more rows of dummy capacitors, the CMOS devices in all amplifier and bias circuits above the C-DAC array, and comparator and logic below the array have uniform spacing and match the capacitor pitch exactly in *X*. Uniform gradients in *Y* will not substantially degrade matching.

8 Interconnect Physical Design

Major foundries are recommending that minimum-design-rule spacing not be used for metal interconnect in advanced technology nodes, except in cases where the circuit area premium is worth the cost in potential yield degradation. Local and long distance interconnect parasitic capacitance Cp and resistance Rp is increasing significantly beyond the 100 nm node as both metal and intermetal dielectric layers become thinner. At the 45 nm node, it is not unusual to find 20% bandwidth differences between pre and post-layout simulation results.

8.1 Clock and Shielded Analog Signal Distribution Physical Design

Figure 10 shows a comparison between shielded interconnects at the 100 nm and 45 nm node. In this example a clock signal is carried on M2, with shields above and

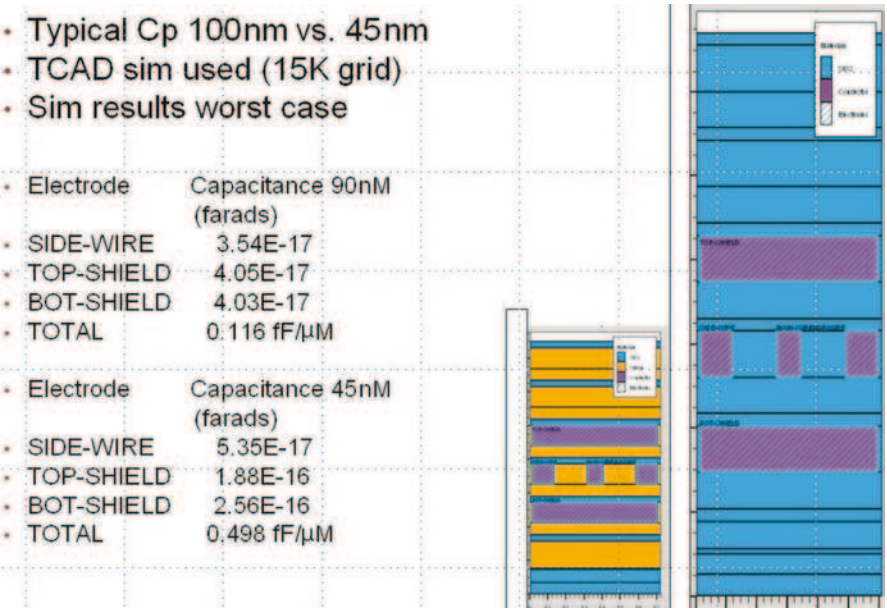


Fig. 10 A scaled comparison of POLY-M2-M4 interconnect dimensions at 100 and 45 nm. The Cp results are obtained using TCAD simulations

below on M4 and poly, respectively. The figure shows the relative scale of the conductor width and intermetal insulator thickness. It is therefore not surprising that a TCAD simulation of the routing channel having the same proportions reveals a Cp/μm more than a factor of 3 larger at 45 nm than at 100 nm.

8.2 Dense Metal Interconnect Physical Design

As dense metal interconnect becomes more dependent on the proximity to adjacent metal patterns, reliability problems from necking, pinching, or bridging effects are becoming more severe. One method for mitigating these problems in the physical design is to increase the regularity of the metal patterns by achieving uniform pitch and using wider-than-minimum widths and spacings.

Several IC design firms are using regular patterns for metal interconnect at the 45 nm node. A representative sample obtained by reverse engineering an Intel 45 nm logic chip [9] is shown in Fig. 11. Note that pattern fidelity is improved by arranging metal patterns primarily in one direction on each layer (in this case M2 and M3).

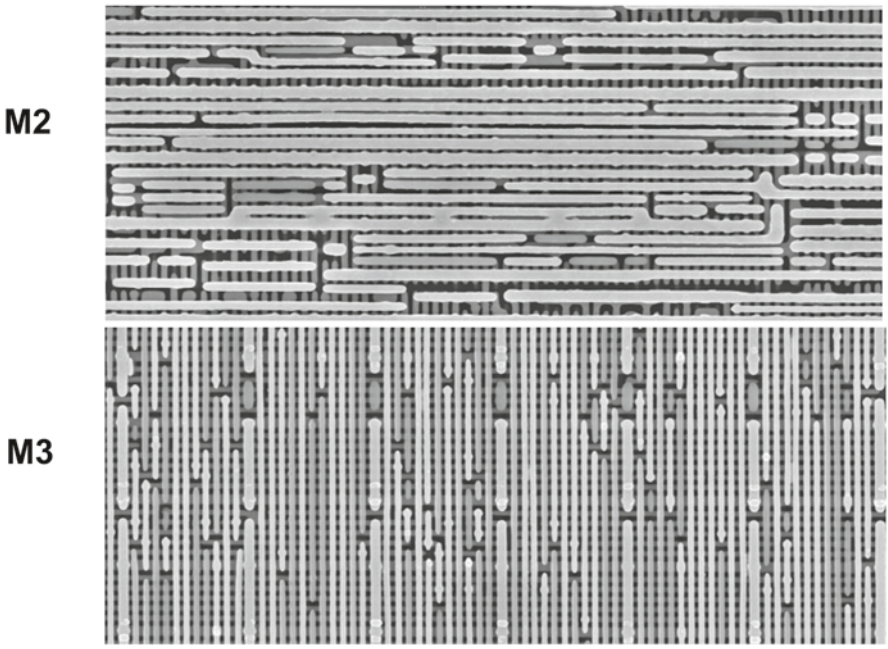


Fig. 11 *M2* and *M3* interconnect patterns obtained by reverse-engineering a 45 nm Intel IC logic area. (Image credit D. James CICC 09)

9 Summary and Conclusions

In the nanoscale era of analog CMOS design, an understanding of the factors affecting physical design is becoming increasingly important. The level of understanding will be reflected in the quality of the initial circuit design. Improper device *W/L* specifications for devices operating at typical analog HF amplifier current densities result in difficulty complying with EM limitations in the physical layout. This is because, in deep nanoscale technology nodes, EM limits constrain maximum HF device *W*. Device physical design uniformity is helpful to assure both lithographic pattern fidelity and simplify the specification of pre-layout STI, WPE, and oxide-definition-spacing simulation parameters. Induced stress from adjacent device patterns and local wiring are becoming dominant physical design considerations.

At the 45 nm node and beyond, there must be close collaboration between circuit and physical design activities in order to assure a manufacturable IC and avoid long rework cycles following post-layout design verification. Where high performance analog circuits are a significant fraction of the SOC, leading IC design companies are requiring analog circuit designers to work more closely with the physical design team, or in some cases, perform the physical circuit layout. The historical separation of circuit and physical design may also be improved by physical designers assuming more responsibility in the circuit design, simulation and verification.

The problems facing analog circuit design in nanoscale technology nodes with severe voltage restrictions and physical design constraints appear formidable. However, they can be mitigated and overcome with the same kind of dedication and innovation demonstrated by Si foundry process development engineers to keep pace with Moore's law in advancing process technology generations every 2 years. Implementing regular device patterns and other physical design strategies at the beginning of a project may require some additional design time. However, these strategies have been implemented on previously-successful IC design projects. The extra planning and implementation time required is small compared to the re-design time required in projects where many physical design defects are found late in the pre-tapeout verification cycle.

References

1. S. Saxena, C. Hess, H. Karbasi, et al., Variation in transistor performance and leakage in nanometer-scale technologies. IEEE Trans. Electron Devices **55**(1), 131–144 (2008)
2. M. Campbell, V. Gerousis, J. Hogan, J. Kibarian, L. Lanza, W. Ng, D. Pramanik, A. Strojwas, M. Templeton, When IC yield missed the target, who is at fault? Panel Session, in *Proceedings ACE IEEE DAC*, 2004, p. 80
3. L. Lewyn, T. Ytterdal, C. Wulff, K.K. Martin, Analog circuit design in nanoscale CMOS technologies, Proc. IEEE **97**(10), 1687–1714 (2009)
4. Y.-M. Sheu, et al., Modeling the well-edge proximity effect in highly scaled MOS-FETs. IEEE Trans. Electron Devices **53**(11), 2792–2798 (2006)
5. C.-Y. Chan, Y.-S. Lin, Y.-C. Huang, S. Hsu, Y.-Z. Juang, Impact of STI effect on flicker noise in 0.13-μm RF nMOSFETs. IEEE Trans. Electron Devices **54**(12), 3383–3392 (2007)

6. R. Klein, Overview of process variability, in *Proceedings ISSCC 2008 Microprocessor Forum F6: Transistor Variability in Nanometer-Scale Technologies*, Feb. 7, 2008, pp. A1–A24
7. M. McNutt, S. LeMarquis, J. Dunkley, Systematic capacitance matching errors and corrective layout procedures. IEEE J. Solid-State Circuits **29**(5), 611–616 (1994)
8. L. Lewyn, M. Loose, A 1.5 mW 16b ADC with improved segmentation and centroiding algorithms and litho-friendly physical design, in *Proceedings IEEE CICC*, 2009, pp. 169, 170
9. D. James, Design-for-manufacturing features in nanometer logic processes—a reverse engineering perspective, in *Proceedings IEEE CICC*, 2009, pp. 207–210

Robust Design for High Temperature and High Voltage Applications

Ovidiu Vermesan, Edgard Laes, Marco Ottella, Mamun Jamal, Jan Kubik, Kafil M. Razeeb, Reiner John, Harald Gall, Massimo Abrate, Nicolas Cordero and Jan Vcelak

1 Introduction

The market segments in which electrical cars will first have significant market shares will range at the beginning from micro/personal vehicles (Class μ: 10–20 kW) up to small personal/family vehicles (Class A: 15–30 kW) and later to family vehicles (Class B: 25–50 kW) [1, 2]. Higher segments will require further developments in storage systems thus being expected to come on the market late 2020–2025 where commercial delivery vans, trucks, minibuses and urban buses together with electric bicycles and scooters have already gained significant market shares: in particular in the Far East the latter are currently overcoming the ICE based ones. The nanoelectronics technologies and power electronics devices, circuits and modules will address the needs of μ, A and B segments. This can be achieved by smart integrated power electronic modules by applying high temperature power electronics and ultra high power density mechatronics.

The architecture of the EVs and HEVs aggregate devices, power modules, processing units, embedded systems, algorithms, mechatronics modules and mechanical parts in five main functional domains [3]:

- Energy (batteries, super capacitors, range extender, grid connection)
- Propulsion (power converters and motor-generators)
- Power and signal distribution (wiring, harnesses and intra vehicle communication)
- Chassis (steering, brake, suspensions and correlated functions)
- Body and board control (HMI, vehicle entertainment, navigation, vehicle to vehicle—V2V—and vehicle to infrastructure—V2I—communication).

The development of power electronics technologies, devices, circuits and modules for these functional domains can be clustered in four design areas:

- Power conversion—AC–DC, DC–DC, high power modules
- Power and energy management—Smart battery management, super capacitors, alternative energy sources and e-grid integration (including power conversion)

O. Vermesan (✉)
SINTEF, Forskningsvn. 1, P.O. Box 124 Blindern, 0314 Oslo, Norway

H. Casier et al. (eds.), *Analog Circuit Design,*
DOI 10.1007/978-94-007-0391-9_4, © Springer Science+Business Media B.V. 2011

- Power distribution network—Power switches, high current sensors and safety fail mode switches
- Smart dynamic monitoring—Information systems and feedback loops based on Si sensors

In the near future the major challenges for the EVs developments are therefore:

- Reasonable driving range (mileage dependent on the terrain profile and navigation)
- Fulfilment of compliance with safety requirements
- Light weight and compact battery system
- Affordable system costs in relation to the operation costs
- Acceptable lifetime of components, intrinsically longer then ICE due to less components, and electric motors that are robust and easy to replace.
- Quick battery recharging

For EVs and HEVs automotive applications, there is a growing demand for smart power solutions integrating power electronics modules with a variety of sensor and actuator functions (Fig. 1) which can be placed directly at the load, thus reducing wiring costs [4].

These solutions are used to process sensor signals, drive actuators and enable further reduction of energy consumption, driver assistance systems and comfort functions. The need for power electronics modules and mixed signal smart power IC's that combine digital and analog signal processing and power switches is increasing. Battery monitoring systems for the electrical vehicles are designed using the smart power technology.

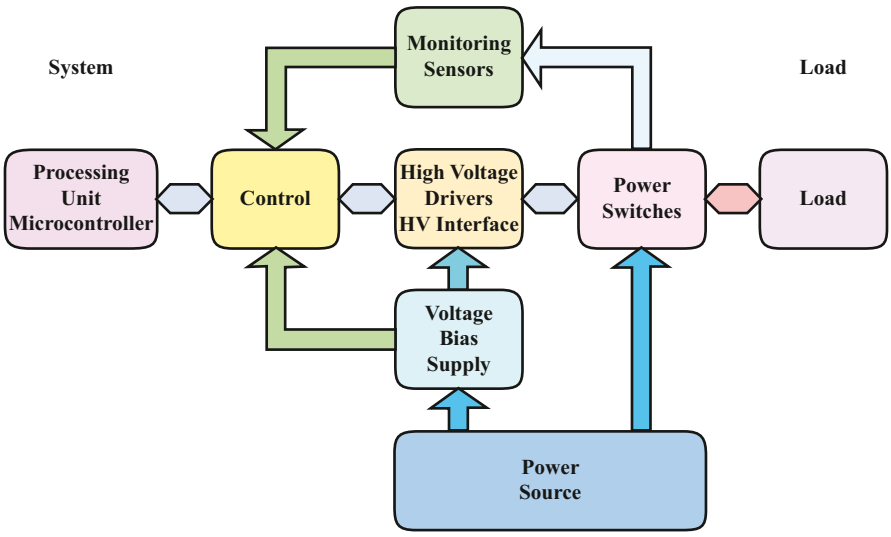


Fig. 1 Function partitioning in a power device

Due to the high operating temperature requirements (e.g. 200°C) and the very stringent quality needs, the current design rules and design practices based on industrial product segments, are no longer sufficient to fulfil the specifications under all operating conditions and achieve zero defects in the automotive environment and advanced tools are necessary to assist the circuit designer in the design for robustness. For example electro thermal simulation provides solutions for determining the temperature of each component on the circuit for all operating conditions allowing to verify circuit performance and to guarantee a lifetime well beyond specification.

The reliability of the semiconductor devices used in EVs and HEVs automotive applications is a very important issue, as these devices and modules have to operate under harsh environmental conditions such as high temperatures, humidity and vibrations. A significant increase in the system reliability can be achieved by monolithic integration of different functions on the same chip, where hybrid solutions typically suffer from reduced reliability owing to many interconnects, plugs and complex mounting procedures.

In this context the power electronics contribute to energy saving and improved efficiency in several aspects:

- Increasing efficiency of power electronic components, modules and systems by optimised component technology or circuit topology e.g. for inverters or power supplies.
- Improving energy efficiency on system level applying power electronics e.g. by introducing variable speed drives or enabling energy recovery or recuperation of electric drives e.g. in HEVs and EVs.
- Energy savings with intelligent power electronics. The key for market acceptance and market penetration is costs requiring a (mechatronic) integration of power electronics in the application system, together with sensors and information and communication technologies.
- Increased operating temperatures, higher packaging densities and increased robustness and reliability.

The new developments in automotive industry on EVs and HEVs have specific requirements to power semiconductors. The operating ambient temperatures can range from 85 to 175°C which requires in turn to the power semiconductor devices to operate at junction temperatures of 175–200°C and over, under maximum currents loads. In addition the standards require electronic devices to work up to −40°C: vehicles need to operate both in very low temperature areas and in very hot regions, thus requiring such reliability requirements. To reach smaller form-factor requirements, temperature and power management are key issues for power and control electronics in vehicles.

To minimise the cooling demand, it would be necessary to increase the maximum allowable junction temperature. For IGBT rated at 600 V, the allowable junction temperature has been raised by 25–150°C operational and 175°C maximum in the last years. This increase can be fully exploited as the conduction losses are almost independent of the temperature and the switching losses are

Table 1 Table Requirements of high temperature electronics

	Sensor and low power	Hydraulic valves and small electric motors	Electric motors	Drives for electric vehicles
Current capacity (A)	<0.5	2–5	20–50	100–400
Power loss (W)	<2	5–15	20–100	>100
Ambient temperature (°C)	130–180	130	100	80
Integration technique	Monolithic		Multi chip	
Packaging	Low cost		Power module	
Development trend	Reduction of the cooling effort R_{thja} from 2 to 10 KW^{-1}		Integration of protective and diagnostic functions at the power device and implementation of the driver IC into the power module	

only slightly increased for higher temperatures. During switching in short circuit tests, some IGBTs showed good robustness at junction temperatures around 200°C.

The motivation to develop high temperature smart power products is presented Table 1 [4] and 2 [1], which shows a summary of the requirements for high temperature electronics. For the packages for multi chip power modules, a junction case thermal resistance of few $K \cdot W^{-1}$ can be achieved by mounting the power devices on proper carriers with the devices specified up to 175°C.

An increase in the maximum chip temperature to 200°C would make system design easier and allow a reduction in system costs by minimizing the required cool-

Table 2 Requirement of power semiconductor devices in EVs and HEVs

Module	Peak power (kW)	Devices	Current (A)	Voltage (V)	Switching frequency (kHz)
Inverter for propulsion	10–100	IGBTs diodes	100–600	200–1200	10–25
Inverter for range extender	40–100	IGBTs diodes	400–600	200–1200	10–25
DC–DC voltage boost converters for battery	10–100	IGBTs diodes	100–600	100–1200	10–25
Heating, ventilating, and air conditioning (HVAC) converters	2–4	IGBT-MOSFETs diodes	10–20	100–400	10–25
Main power DC–DC converters for 14 V auxiliaries	1–2	MOSFETs diodes	20–40	200–400	50–200

ing effort. In this way the packaging becomes cheaper, the power module becomes easier to build, and the chip area of the semiconductor device can be made smaller.

Smart power applications are spread over a broad voltage range. Similar requirements for high temperature applications are found in the power electronics of different transportation systems such as cars, airplanes, satellites, and even railway systems. The market for high temperature high voltage devices spans from low voltage automotive applications (up to 60 V) to grid connection (110–220–380 V AC) up to electronics for power traction (400 V and over).

2 Nanoelectronics Semiconductor Technologies for Power Devices for EVs and HEVs Applications

High temperature power electronics is therefore the main enabling technology for the forthcoming electrical mobility: In EVs and HEVs bi-directional power inverters (DC–AC) and converters (DC–DC) have to manage some tens of kilo-volt-amperes with power losses in the order of 3–8% and thus generating several hundreds of watts (and up to some thousands) of wasted heat.

In EVs and HEVs applications, power MOSFETs are generally used for systems with less than 100 V DC and IGBTs (Insulated Gate Bipolar Transistors) are used in systems requiring voltages over 200 V DC. In addition in some applications ranging from around 100–200 V DC, MOSFETS are still used for low power applications and IGBTs are used for high current applications. Bipolar Darlington power transistors, MOS-Controlled Thyristors (MCTs), and thyristors are not used in automotive applications. The technological trends for power devices are presented in Fig. 2.

The trends in the IGBT module on increasing their power capacity greatly help to contribute to system downsizing. The optimum combination of the new silicon and new package technologies results the most economical solution for each module ratings.

SiC unipolar devices (MOSFETs) have a potential to replace bipolar IGBT devices in various application fields, including automotive power electronics and very high voltage systems; however several critical issues remain, like improving wafer size and quality, solving device processing hurdles …

3 Robustness, Reliability, Dependability

In propulsion applications where the vehicle is subject to frequent acceleration and deceleration, the power devices are in turn subject to strong and quick thermal stresses. Similar stresses may occur for other auxiliary motor drives or DC–DC conversion applications.

The power semiconductors and packages must show very high reliability: components must comply with Automotive Electronics Council requirement AEC-Q101 industry standard for power discrete and smart power based integrated circuits.

Fig. 2 Technological trends for power devices [1]

Power semiconductors must pass 175°C High Temperature Reverse Bias (HTRB) and 175°C High Temperature Gate Bias (HTGB) testing. The temperature requirement is increased to 200°C to ensure that the semiconductor devices are robust and reliable in terms of latent silicon defects that can be generated during the wafer processing fabrication.

The reliability assurance requires the evaluation of parameters from technology, system and stress profile in order to generate the stress on the specified electric vehicle module and system in operation. Based on the system description and the system simulation (thermal, corners, etc.), reliability risks can be derived from the results and corresponding accelerated tests are developed to a particular electronics module in a specific electric vehicle environment. The concept is presented in Fig. 3.

The challenges for the smart integrated power modules for the HEVs and EVs are:

- Power efficient and loss reduction for semiconductor technologies
- Plug and play standardised solutions

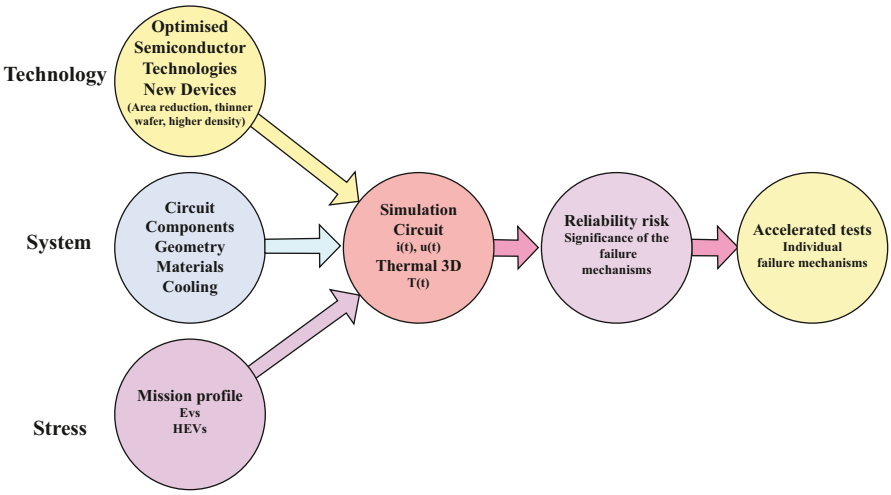


Fig. 3 Methodology for reliability evaluation

- Cost effective packaging solutions
- Highest reliability and optimum design
- Increased thermal performance
- Improved thermal cycling capability (lifetime, ppm)
- Higher power density
- Smallest possible footprint

During the system definition of the power electronics system it is important to ensure that the modules are cost effective and compact and are implemented based on the following design principles:

- Use of components that are unpackaged (bare multi chip modules)
- Minimise the length of the electrical connections
- Minimise the inductance of the interconnections
- Cool optimally the components (coolant temperature from −40 to 125°C; normal temperature 90°C)
- Use only one hermetically sealed overall housing

Designing the electronics modules for EVs and HEVs requires following the steps listed below:

- Definition of circuit topology
- Selection of components and the associated compact models (Spice, etc.)
- Determine the minimal geometry of the electrical interconnections
- Calculation of the losses in the components
- Layout of the cooling (thermal paths).
- Thermal simulations and determination of the maximum temperature stress on the components

- EMC analysis
- Implementation of demonstrators/prototypes

The elements that have to be considered during the design process are:

- Materials: thermal conductivity, coefficient of thermal expansion CTE, machining properties,
- Active and passive components: power losses; safe operating area (SOA);
- Connections:
  - Solders (melting points, creep characteristics, wetting; machining temperature)
  - Bond wires (cross section; losses)
  - Interconnections (cross section, losses)
- Cooler: cooling capacity for a specified flow rate
- Sensors: current, voltage and temperature measurement
- Mission profile: the way in which the traction stress affects the system can be derived from the traction cycles. The important input parameters are the current paths required by the load.

In the electronics modules the reliability risks can be identified by using the failure mode cause and effect analysis into four areas:

- Components
- Connections (of the devices and components)
- Package
- Operation (of EVs and HEVs).

The failures are evaluated based on following criteria:

- Occurrence
- Significance
- Discovery

In many cases one of the most frequent causes of failure in electronic systems comes from the dissimilar coefficients of thermal expansion of the materials used.

In order to determine the physics of failures for the electronics for EVs and HEs different type of test are required. Accelerated tests must be developed and verified for each specific fault mechanism. For example, a highly accelerated test is applied to bond wires. The fault mechanism can have two elements, the different coefficients of thermal expansion of the bond wires compared with the metallization on the silicon chip, and the natural oxide layer formed on the solder.

System dependability, defined as the quality of a delivered service such that reliance can justifiably be placed on this service, is evaluated by measuring the following magnitudes:

- Reliability: the ability of a system or component to perform its required functions under stated conditions for a specified period of time,
- Availability: the proportion of time a system is in a functioning condition;
- Safety: the freedom from unacceptable risk of physical injury or of damage to the health of people either directly or indirectly (through damage to property or to the environment).

For the electronics modules used in EVs and HEVs the failure is considered the event that occurs when the delivered service of the electronic system deviates from the specified mission. In this case the error is the part of the system state which is liable to lead to failure and the fault is the phenomenological cause of the error. The most frequently used reliability measure for semiconductor devices is the failure rate (λ). The failure rate is obtained by dividing the number of failures observed, by the product of the number of devices observed and the number of hours operated, usually expressed as the percent of failures per thousand hours, or failures per billion device hours (FITS).

The semiconductor industry uses highly accelerated testing procedures to assess the reliability of semiconductors. During accelerated tests, combinations of stresses over the normal values, are used to produce, in a short period, the same failure mechanisms as would be observed during normal use conditions. Temperature, relative humidity, and voltage are the most commonly used stresses used during accelerated testing. For instance the devices are tested near their maximum junction temperatures and failures are analyzed to verify that they are of a type that is accelerated by temperature. This gives an estimation of most probable failures, but does not give any information on when the failure will occur. Afterwards ad-hoc tests over the most probable failure mechanisms, together with combination models, are used to estimate the failure rate λ.

In order to design a dependable system a combination of the following techniques are considered for the electronics modules:

- Fault avoidance, to avoid faults by design;
- Fault removal, to reduce by verification the presence of faults. Fault injection techniques are fundamental in this method;
- Fault tolerance, i.e. techniques that are able to detect the faults or provide correct operation despite presence of faults;
- Fault evasion, to estimate by evaluation the presence, the creation and the consequences of faults and to take pre-emptive actions to stop the fault from occurring;
- Fault decontamination to eliminate or reduce the effect of a fault once this occurs in a system.

A combination of these techniques must be used to obtain a dependable system.

In the electronic modules for EVs and HEVs the complexity of the systems requires a system level approach in order to cope with all these aspects and to achieve the best compromise between reliability and costs; moreover, it is only a combined utilization of different techniques that can lead to a dependable system. In this context it is necessary to decrease detection and correction costs, and move the focus to failures instead of faults.

In the case of electronics modules for EVs and HEVs it is necessary to use a methodology including simulation software reliability analysis, use of fault injection at different abstraction level and selection of a mix of hardware and software protecting techniques and thus creating a fault robust approach for high temperature and high voltage applications, where robustness is considered the ability of a module to continue to function despite the existence of faults in its component subsys-

tems or components, even if system performance may be diminished, but always in a reliable way, until the faults are corrected.

The goal of robust design for electronics modules for EVs and HEVs is to create circuits whose characteristics remain within the called design requirements regardless of possible semiconductor manufacturing process variations and changing operating conditions.

A circuit characteristic is any real quantity that can be measured on a circuit. The operating conditions (i.e. temperature, power supply voltage, etc.) are bounded and the bounds are agreed upon before the designer starts the design process. The manufacturing process variations which can occur during IC fabrication are provided by the circuit foundry. We focus on those design problems where these variations are provided in the form of corner models of circuit elements. For a CMOS process usually the following corners are provided: worst one (WO), worst zero (WZ), worst power (WP), worst speed (WS), and typical mean (TM). Corner models can also be provided for resistors, capacitors, IGBTs, etc.

In order to obtain a robust circuit an additional step of design centering is required by using statistical or deterministic design centring techniques.

Robust design used by IC designers relies on the assumption that circuit characteristics have their extreme values at the extremes of operating conditions and process variations. A circuit is checked to be robust by simulating it against the combinations of extreme operating conditions and process variations. This assumption is valid only if circuit characteristics are monotonic functions on the intervals enclosing operating conditions and model parameters of corner models.

3.1 Safe Operating Area

For the electronic modules for EVs and HEVs the total safe operating area (SOA) contains:

- Electrical SOA: the Ids–Vds limit for the activation of the intrinsic bipolar transistor by electrical effects. This is especially important for very short pulses (up to a few hundredths of nanoseconds) like in ESD events.
- Hot carrier SOA: limiting the device use due to the slowly degrading transistor characteristics as a result of interface trap generation and carrier trapping. The hot carrier effect plays on a time scale of seconds to years.
- Thermal SOA: the Ids–Vds limit for the activation of the intrinsic bipolar transistor by thermal effects. Pulse times are typically in the sec.-msec. range, like e.g. in inductive switching.

Electronics modules for EVs and HEVS power semiconductor IGBTs and MOSFETs, operate in power inverters to control the flow of energy from the battery to the motor. They are placed in the circuit in between the accumulators and the AC motor. The flow of energy is controlled by commanding the power devices to switch on and off repeatedly. The Safe Operating Area for a power semiconductor IGBT or MOSFET is the maximum operating condition which must be met during the switching transients. This condition is a function of the simultaneous voltage

imposed across the device and current flow through it. These transients may be described as turn ON and turn OFF, and each of these transients may have a characteristic stress on the power semiconductor device.

The SOA of a device is an absolute rating and the failure mode when exceeding it, is a catastrophic failure of the device. To get highest reliabilities, the SOA of a device is defined in such a way to allow the device to handle the worst case transient current simultaneously with the highest input voltage without damage.

4 Simulation

4.1 Electro-Thermal Simulation

Integrated power devices such as DMOS transistors can generate a significant amount of heat when switching high currents (several Amp's).

Thermal simulation has to consider the circuit in operation with the heat flowing away from the heat source and increasing temperature of the neighbouring devices; this simulation is iterative since the transistor currents are temperature dependent. In this simulation, all aspects have to be taken into account: metal resistance is not negligible at the high currents and the metal resistance is temperature dependent as well.

Currents in large drivers might not be homogeneous and the simulation needs to consider the layout of the switch, including all the resistors. The thermal resistance of the package is another key parameter.

The Heatwave package is commonly used for dynamic electro-thermal simulations: the outcome of the simulation is a temperature profile at each location of the die during the circuit operation. It flags to the designer the local hot spots and the associated risks for thermal run-away. It must also be noted that temperature gradients over the die could lead to other artefacts such as mismatch of transistor currents and parameter shifts.

4.2 Reliability Prediction by Ageing Simulation

Stressing the devices by voltage or current can lead to parameter shifts or even failures such as metal opens. It is well known that these risks are strongly temperature dependent (typically exponential).

A model of the device ageing (parameter shifts) as a function of applied voltage, current and temperature can be generated based on extensive device stressing experiments. Possible ageing and failure modes for the silicon devices are hot carrier instability of transistors, time dependent dielectric breakdown of gate oxides and electro-migration of metal interconnections on chip [5].

Failures related to the package are even more difficult to model since they not only depend on electrical and thermal stress but also on mechanical stress in the

package. Modelling requires the characterization of the package moulding compound during thermal cycling.

Basic models are typically implemented in quick and dirty reliability calculators (running in EXCEL or MathCAD environments) to be used by the designers for a first dimensioning design. Each calculator typical considers one reliability aspect at a time (e.g. electro-migration of metal lines).

Advanced reliability simulators considering several ageing aspects together are used as a final design check over a block or a complete circuit. The outcome of the reliability simulator can be a maximum parameter shift for a certain transistor at the end of life or flags that indicate that certain devices or blocks poses some risk and should be further analysed.

However both quick and dirty calculators and reliability simulators can handle complicated temperature profiles e.g. 15 years at 150°C accumulated with shorter times at respectively 170, 180,…, 200°C.

5 Packaging and Interconnects

Multi chip power modules are predominantly used in EVs and HEVs. The hybrid modules distribute signal and power, dissipate heat, protect the devices enclosed, and serve as the basic power electronics building block.

Power module design has to address the inherent mechanical stresses after bonding of large silicon chips having low coefficient of thermal expansion (CTE) with other materials having higher CTE. Automotive modules must function down to −40ºC where CTE mismatch stresses may become excessive.

The power module technology relies on the use of aluminium wire bonds, direct bond copper ceramic substrates, and copper base plates. The thin aluminium wire bonds suffer from high parasitic impedance, fatigue-induced lift-off failures, and inability to remove heat.

The DBC ceramic substrate (Al_2O_3 or more expensive AlN) provides electrical isolation but inadvertently increases the package thermal resistance. The thick Cu base plate serves as a heat spreader but considerably increases the weight, size, and thermal resistance of the power module.

Power semiconductor modules often contain several MOSFETs/IGBTs and diodes, which are soldered onto a metal plated ceramic substrate. To connect the top side of the chips, wire bonding is usually employed. Multiple substrates are connected to a base plate by use of soft-solder joints.

Changing the maximum allowable junction temperature of the power semiconductor will directly change the thermal stress on the interconnection of the chip surface. A typical wear out effect at the chip surface is the wire bond lift off. To test this interconnection, power cycling tests are performed.

The number of cycles that a device survives is related to the temperature swing, the maximum temperature and the slopes. For the introduction of a maximum junction temperature of 175°C, the wire bonding process requires to be improved.

Over the lifetime of direct bond copper (DCB) modules, the layers are prone to recurring mechanical stress, due to the ongoing thermal cycles.

Caused by the current flow in the semiconductor and the resulting heat-up, the materials used, such as copper, ceramics, silicon and aluminium expand with their different coefficients of expansion.

Many power module solutions make use of Al_2O_3 DCB in conjunction with a copper base plate. This combination of materials is mainly suitable for EV systems and some full hybrid systems.

When designing the power semiconductor module, particular consideration needs to be given to the load profile during the lifetime of the EV and HEV. Once the required profiles are available detailed in passive temperature fluctuations and current profiles, a suitable combination of materials (DCB/base plate) can be determined.

5.1 Metal Carbon Nanotube Composites for High Temperature Applications

Excessive heat can degrade the performance, life and reliability of electronic components used in power semiconductor devices, where removal of this heat is crucial. On that basis, thermal interface material and the metallization poses the major role to heat removal in the system. Carbon nanotubes (CNTs) are allotropes of carbon having novel properties, for example unique electrical and thermal properties that make them useful in various applications such as nanotechnology, electronics, optics and other fields of materials science. The electric current carrying capability of Single-wall carbon nanotubes A cm^{-2} (SWCNTs) is estimated to be 109 A cm^{-2} while copper wires burn out at about 106 A cm^{-2}. More over, the thermal conductivity of these nanotubes is 3,000 Wm^{-1} K^{-1} compared to 401 Wm^{-1} K^{-1} for copper.

Composite materials made of metal and CNT are expected to provide unique electrical and thermal interface in the condition of high voltage and in high ambient temperature. Here we have dispersed CNT into Nafion, polyacrylic acid and chlorosulphonic acid in the presence of Ag/Cu/Ni, followed by electrodeposition onto Au sputtered Si substrate. It has been characterized using X-ray diffracting and carbon analysis to confirm the metal and the amount of carbon percentage.

Surface morphology has been examined using scanning electron microscopy (SEM) and electrical resistivity is measured using standard four point probe method, where electrical resistivity found to be varied with the deposition condition of the composites (Fig. 4).

5.2 Thermal Management

At high junction temperatures both the performance and the long term reliability of electronic devices rapidly degrades. In power devices operating at high ambient

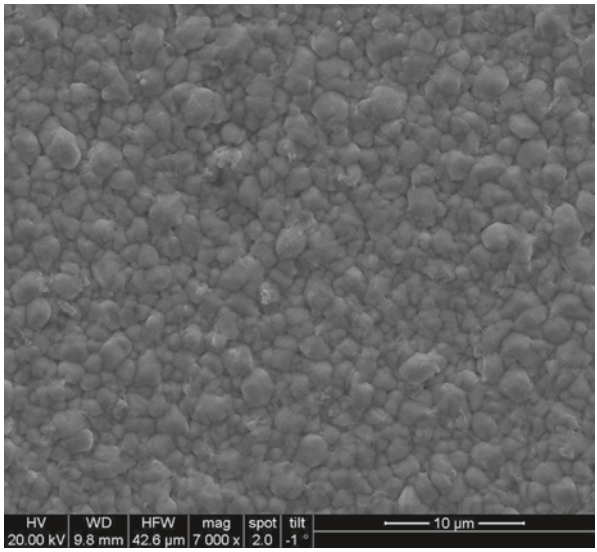


Fig. 4 SEM image of the electrodeposited Ag-CNT composite on Si

temperatures, the heat dissipation is higher and the temperature margin (difference between maximum junction temperature and ambient temperature) narrows. Therefore a global thermal management solution optimised at all levels (device, module and system) is required to achieve a lower junction temperature.

At device level, the thermal management solution will provide a good thermal path for heat spreading and heat transport away from the junction.

High conductivity materials and thermal interfaces are therefore essential. At module/package level, a good thermal path from the chips to the board is required, which meanwhile does not introduce additional stresses which would compromise the mechanical reliability of the module.

Hence, at module and mostly at device level, heat conduction is the most important factor to optimise. Finally, at system level, the heat which had been carried away from the heat sources must be dissipated to the surrounding ambient by convection (and sometimes radiation). Passive (e.g. heat sink) or active (e.g. fan) solutions are required. However, an overall optimised thermal management will only be achieved if all levels are considered jointly.

Computational Fluid Dynamics (CFD) solvers are an essential tool to support the design and optimisation of thermal management at all levels, as they can accurately predict the conduction, convection and radiation behaviour.

6 Conclusions

Modern automotive circuits have to operate at higher temperatures and need very high quality levels (target to zero defects). New tools such as thermal electric simulators and reliability simulator assist the designer to achieve these goals for high robustness.

Thermal electric simulators are fully deployed by semiconductor companies for the design of automotive smart power circuits. Reliability calculators are in use for many reliability aspects while the use of the novel multi-physic reliability simulators is in an early stage.

Thermal simulations can be used to adjust the thermal characteristics of the device to the worst case operating conditions and the maximum temperature allowed.

The key to optimal smart/intelligent power modules design for HEVs and EVs is to maximise the compatibility of the different technologies (semiconductor, packaging, materials, etc.) to deliver a high reliability modular and standardised modules that are robust and operates at higher temperatures. Planar and stacked configurations will be used for the integrated power modules and as the requirements for miniaturisation are becoming critical, the stacked substrates approach may become preferred.

Acknowledgment The work was supported by the ENIAC E^{3}Car project under the project grant. The authors would also like to convey thanks to the Public Authorities and European Commission for providing the financial support. Reliability simulation research is done by ON Semiconductor in the frame of the MEDEA+ELIAS project funded by IWT in Belgium.

References

1. O. Vermesan, R. John, M. Ottella, H. Gall, R. Bayerer, High temperature nanoelectronics for electrical and hybrid vehicles, in *Proceedings of IMAPS International Conference on High Temperature Electronics Network* (*HiTEN*), Oxford, UK, 13–16 Sept 2009, pp. 209–218
2. R. John, O. Vermesan, R. Bayerer, High temperature power electronics IGBT modules for electrical and hybrid vehicles, in *Proceedings of IMAPS International Conference on High Temperature Electronics Network* (*HiTEN*), Oxford, UK, 13–16 Sept 2009, pp. 199–204
3. S.A. Rogers, *2008 Annual Progress Report for the Advanced Power Electronics and Electric Machinery Technology Area* (U.S. Dep. Energy, Jan 2009)
4. J. Korec, Silicon-on-insulator technology for high-temperature, smart-power applications. Mater. Sci. Eng. B. **29**(1–3), 1–6 (1995)
5. P. Moens, G. van den Bosch, Reliability assessment of integrated power transistors: Lateral DMOS versus vertical DMOS. Microelectron. Reliab. **48**, 1300–1305 (2008)

Radiation Effects and Hardening by Design in CMOS Technologies

Federico Faccio

1 Introduction

Radiation threatens the correct functionality of electronics in space, avionics, nuclear and High Energy Physics (HEP) applications. The engineering community involved in each of these disciplines has developed over time its specific set of solutions to ensure system reliability. The specificity is dictated by requirements in terms of cost, accessibility and mission criticality of each component/system. In some cases, the selection and use of Commercial-Off-The-Shelf components (COTS) represents the cheapest and most practical solution. Nevertheless, the need for extensive irradiation campaigns in view of selecting first, then assuring reproducible radiation response of the components increases the cost and required testing resources. On the other hand, the development or purchase of dedicated "radiation-tolerant" components eases qualification and procurement and ensures the required level of reliability, but is typically more expensive. A small catalog of radiation-tolerant components is in fact accessible commercially at a cost/function ratio considerably larger than for COTS. This cost is in part determined by the need for using a dedicated manufacturing process (most often CMOS) qualified against radiation effects. The low volume of this market, together with its strict quality assurance requirements, make the business profitable for a small number of "niche" foundries only if silicon wafers can be sold at large prices for a relatively long time. Processes are therefore not phased out as quickly as in the commercial marketplace, and new developments take more time. As a result, they are typically lagging two generations or more behind commercial-grade products.

An alternative approach, increasingly popular in the last decade, is based on the use of commercial standard CMOS technologies. Integrated Circuits (ICs) are designed with dedicated techniques to withstand radiation effects, an approach that is called "Hardness By Design" (HBD). This allows circuit designers to using state-of-the-art processes to design ICs, while at the same time delivering radiation-tolerant components. In doing so, designers must be aware of the radiation effects

F. Faccio (✉)
PH dept., CERN, 1211 Geneva 23, Switzerland

H. Casier et al. (eds.), *Analog Circuit Design,*

DOI 10.1007/978-94-007-0391-9_5, © Springer Science+Business Media B.V. 2011

against which circuits have to be protected. For this reason this paper, which is aimed at summarizing the field of HBD in CMOS technologies, starts by illustrating how radiation affects modern CMOS devices.

2 Radiation Effects in Modern CMOS Technologies

Radiation effects can be classified in two main categories: cumulative effects and Single Event Effects (SEE). The slow integration of radiation-induced defects, traceable to either ionization (Total Ionizing Dose, TID) or non-ionization (Non-Ionizing Energy Loss, NIEL) phenomena in the semiconductor where the circuit is manufactured, fall in the first category. SEEs are instead due to the prompt charge deposition originating from a single ionizing particle in the semiconductor, and as such have a stochastic nature: the effect observed depends on the position of the hit, energy and ionization state of the particle, and other momentarily conditions of the circuit.

CMOS technologies are known to be sensitive to TID cumulative effects, but not to displacement damage, which is the main signature of NIEL in silicon. Exceptions to this rule are found in power transistors such as IGBTs and power MOSFETs, and in modern LDMOS transistors as well, which are sensitive to displacement damage [1]. Modern LDMOS are actually degrading their performance only at high particle fluence, much larger than the typical radiation environment in space applications (but not larger than the one found in the latest generation of HEP experiments, or in civilian nuclear applications).

2.1 TID Effects

The CMOS sensitivity to TID effects is traceable to radiation-induced charge trapping in the gate or isolation oxides. Ionizing radiation creates electron-hole pairs in the oxide, with consequent trapping of holes or activation of defect states at the interface between the oxide and silicon (interface states). Depending on the position, function and nature of the oxide where these phenomena happen, the consequences on the circuit can be different. In the gate oxide, the electric characteristics of the transistor might be affected (threshold voltage, mobility, noise). In the thicker lateral STI oxide, hole trapping can introduce an electric field sufficient to invert the lightly doped p regions, opening conductive channels in between n+ diffusions. These leakage currents can flow between source and drain of the same NMOS transistor, or between adjacent unrelated n+ diffusions (including the n-well). Trapped charge can be de-trapped or “annealed” by thermal excitation at a rate dependent on the energy of the traps.

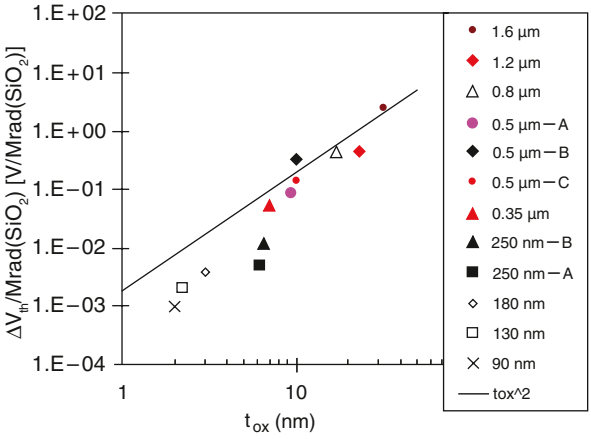


Fig. 1 Threshold voltage shift per 1 Mrad of TID on samples of NMOS transistors in different technology nodes. The points at 180 nm and below are only indicative, since the V_{th} shift is so small that a large statistical study should be done to determine the real value. With the sample size available, the error bars are comparable to the measured values (1–3 mV)

Research work in the 1980s already revealed how the density of trapped holes and interface states both depend on the oxide thickness: the thinner the oxide, the milder the radiation effect [2]. This has been confirmed since by several sources on commercial-grade gate oxides in advanced CMOS. In Fig. 1, the radiation-induced threshold voltage shift for transistors in different technology nodes is plotted versus the respective gate oxide thickness. Already around 5 nm, the shift is practically negligible for most applications: 10 mV or less after a dose of 1 Mrad (which exceeds the requirements in TID for most space missions). At 2 nm, gate oxide thickness typical of the 130–90 nm nodes, the radiation-induced shift is negligible even at doses of the order of 100 Mrad. Additional evidence of the robustness of the gate oxide emerges from noise measurements in these technology nodes: exposure to TID up to 100 Mrad does not have any measurable impact on the noise of the transistors [3]. Unfortunately, no data is available on commercial high-K dielectrics replacing nitrated oxides as gate dielectrics in very advanced CMOS (as from 45 nm). Experimental data is not very useful in this case, since it only concerns laboratory dielectrics which might not be even similar to what actually used industrially.

While the thickness of the gate oxide decreases with down-scaling, the STI oxide thickness remains very comparable. Hole trapping in this oxide—leading to leakage currents—is therefore still a relevant failure mechanism. In particular, the leakage problem is worst at the edges of NMOS transistors where the polysilicon gate extends beyond the transistor channel and overlaps the lateral STI. When the NMOS transistor is turned on, an electric field is also present across the STI at the edge of the transistor, enhancing hole trapping. The accumulation of trapped holes in the STI reinforces the electric field in this region, and eventually inversion at the transistor edges occurs even if the main transistor is turned off, and leakage current can flow in the inversion layer [4].

Contrary to the gate oxide case, where for the same thickness the radiation response of samples from different manufacturers is very comparable (Fig. 1), expe-

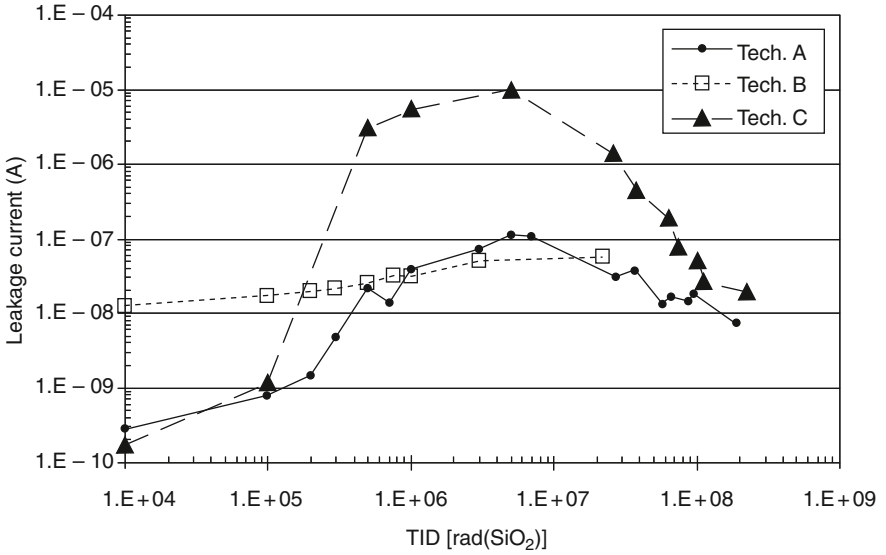


Fig. 2 TID-induced source-drain leakage current for minimum size 130 nm NMOS transistors (≈0.15/0.12 μm) from three different manufacturers

rience reveals a large variability in the radiation response of the STI oxide. As an example, Fig. 2 shows the source-drain leakage current of minimum size NMOS transistors versus TID for samples in the 130 nm generation from three manufacturers [5]. The peaking of the leakage at doses of 1–3 Mrad has been studied in detail and attributed to the compensating contribution of negative charge trapped in defect states at the interface between the STI oxide and the silicon (that compensate for the holes trapped in the STI oxide) [6].

More recent data in 90 nm, shown in Fig. 3, confirm the above observation on the variability of the response. It is interesting to note that the source-drain leakage current evolution is comparable for manufacturers A and B in 130 and 90 nm. This indicates how the FEOL processing details related to the STI and the well/channel doping largely determine the radiation response, which remains comparable for the same manufacturer as long as the same manufacturing "recipe" is maintained in different nodes.

Trapping in the STI oxide can also lead to the opening of conduction channels between n-wells and n+ diffusions, or amongst n+ diffusions at different potential. This possibility is experimentally studied on custom test structures using Field-Oxide FETs (FOXFETs), MOS transistors where the STI oxide is used as gate insulator. Source and drain of the FOXFETs can be n+ diffusions or n-wells, and the gate can be chosen to be polysilicon or even metal. Before exposure to radiation, the threshold voltage of these transistors is typically well above 10 V, and no current flows in the device when the gate and drain are biased at the maximum Vdd (up to 3.3 V for I/O transistors in 130 nm). Figure 4 illustrates how before irradiation the threshold of a FOXFET with n-well drain and n+ diffusion as source is higher than

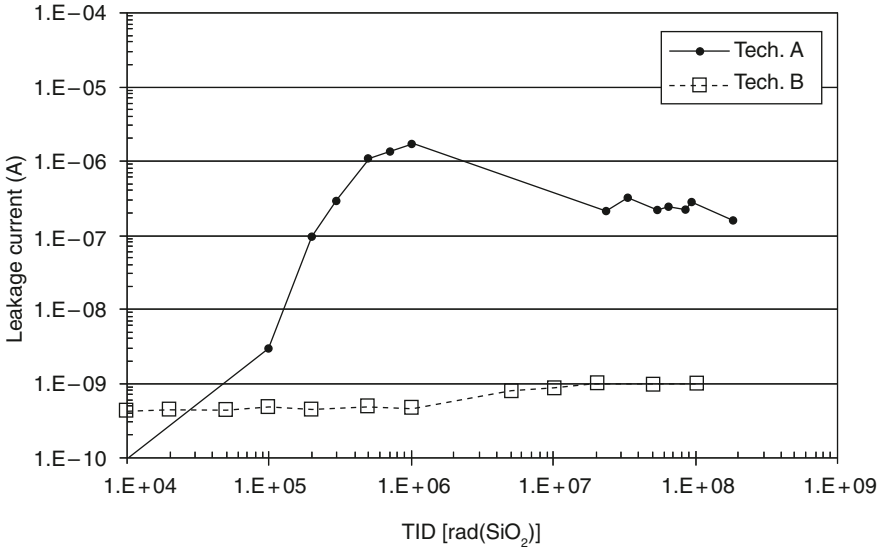


Fig. 3 TID-induced source-drain leakage current for minimum length high-Vt (low power) 90 nm NMOS transistors from two different manufacturers

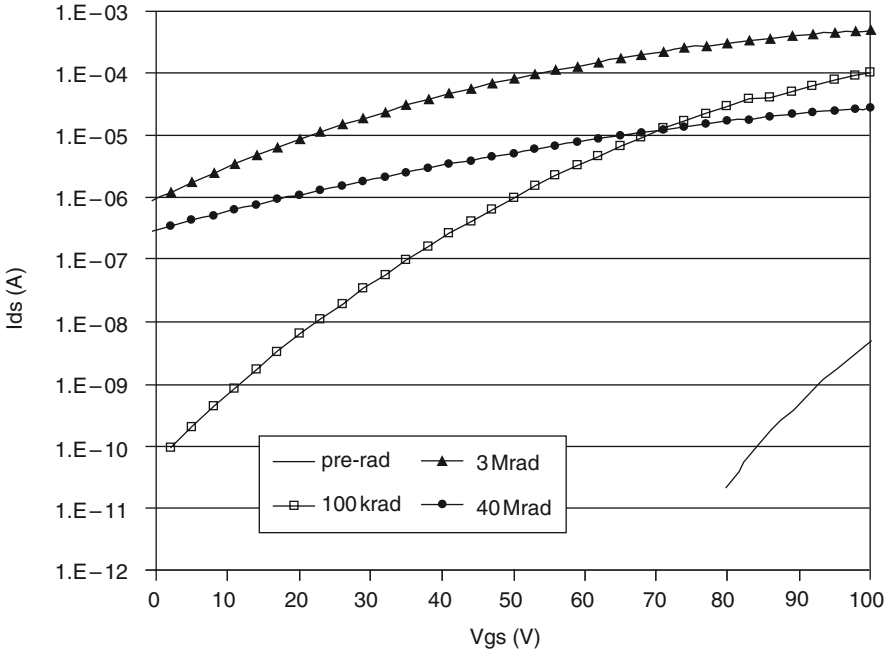


Fig. 4 Current flowing in a FOXFET using n-well as drain and n+ diffusion as source, and poly-silicon gate (W/L = 200/0.03 μm) before and after exposure to TID up to 40 Mrad(SiO_2)

100 V. However, hole trapping in the STI reduces the threshold voltage until current flows in the device also when no bias is applied to the gate—the leakage current in the example reaches 1 μA. The comparison of available data in different technology nodes reveals how this effect gets milder in 130 and 90 nm with respect to older processes: significant leakage is observed only from n-wells. However, these measurements on a small sample of technologies can not be generally extrapolated: different processing details can likely lead to very variable radiation response.

2.2 Single Event Effects (SEEs)

Charge deposited in the ICs by ionization from the transit of an individual particle (heavy ion or proton in space, neutron in avionics and terrestrial applications, proton or neutron or other hadron in HEP applications) can affect the circuit functionality in different ways.

If the hit perturbs the logic state of a memory or register, causing a switch to the "wrong" state, the event is called Single Event Upset (SEU). Its consequences might be unobservable if the stored bit is actually not used, or cause the whole IC to hang in a wrong state requiring a reset—in which case the event is called Functional Interrupt (SEFI). Because of the increased speed of new technologies (the gate delay is shorter than the duration of the perturbation introduced by the strike of the ionizing particle), hits in the combinatorial logic can now propagate through a long series of gates. These glitches can reach the input of a register where, if they hazardously happen to be synchronous to the proper clock transition, they can be latched [7] giving origin to a "Digital" Single Event Transient (DSET). Since the wrong value can only be latched during a clock transition, the DSET error rate depends linearly on the clock frequency.

In analog circuits, Single Event Transients (SET) at any node can be introduced by particle hits. These are especially relevant at the input of amplifiers, in high impedance nodes, and in the presence of precise small currents.

In some cases, SEEs can lead to permanent failure. In low-voltage CMOS this can be traced to the onset of a latch-up, triggered by the current from the particle hit in any internal node of the circuit (not necessarily at the I/Os, where specific structures to prevent electric latch-up are introduced to protect the IC from improper powering sequences). Other destructive events, typical of high power circuits/devices, are Single Event Burnout (SEB) and Single Event Gate Rupture (SEGR), both leading to the destruction of the transistor's gate oxide.

Single Event Upset. Sensitivity to SEU clearly depends on the ratio between the charge stored in a node during normal operation versus the one collected from a particle hit. Since the former decreases with down-scaling (both node capacitance and Vdd decrease), a larger sensitivity to SEU was generally expected for modern technologies. This forecast actually did not materialize. Extensive data has been collected by semiconductor industries to strictly monitor the SEU sensitivity of

memories and CPUs. Such works demonstrated that the sensitivity of individual memory cells has instead been scaling down for both DRAMs, and SRAMs [8]. Recently published data, comparing per-bit sensitivity of memory cells in five consecutive generations from the same manufacturer, clearly show a 20-fold decrease of the error rate when going from 250 to 65 nm [9] (actually, comparison of predicted error rates for the two cells in a geostationary orbit indicates a factor of 60 difference). This highlights how node capacitance and voltage supply alone are not a sufficient indicator: the decrease of the sensitive area and of the charge collection efficiency with down-scaling also plays an important role in the overall SEU response.

The reason why most semiconductor industries, which do not target the niche market of radiation tolerant components, have made the effort to carefully characterize SEU sensitivity of their products lies in the growing concern for large neutron-induced error rates in terrestrial large-scale applications [8]. Radiation is present at ground from natural sources—mainly from cosmic rays—and this radiation background can significantly contribute to increase the Failure In Time (FIT) of the components, especially for large memory banks. In particular, thermal neutrons have been shown to largely dominate the error rate in some components. These very low energy neutrons (around 25 meV) have a large cross section for interaction with an isotope of Boron (^{10}B) whose abundance in natural Boron is close to 20%. The interaction, which takes place mainly in the heavily B-doped BoroPhsophoSilicate Glass (BPSG) layer often used to separate transistors from the BEOL in CMOS, yields ionizing particles (Li and alpha) capable of inducing SEU [10]. When this mechanism was identified, semiconductor manufacturers replaced BPSG with an alternative material, effectively reducing the FIT of their components.

Single Event Latchup. Due to the presence of multiple adjacent n- and p-doped regions in CMOS technologies, parasitic npnp or pnpn structures called thyristors can be found all over the ICs. Such structures can be turned on by the flow of current induced by an ionizing particle's hit. This can lead to a latched low-impedance path between Vdd and ground which, if not promptly interrupted, can permanently damage the device. Because of that, SEL has been a source of concern since the early usage of CMOS components in radiation environments. In the last 15 years, the introduction of STI and retrograde wells together with the steady reduction of the voltage supply has been very beneficial in reducing the general sensitivity of ICs to SEL [11]. However, this alone does not guarantee immunity from latch-up, since SEL is extremely design dependent and very sensitive components can still be found.

Single Event Burnout. SEB is generally associated to power transistors such as IGBTs and VDMOS rated to stand high voltages (300–1,000 V). This destructive event can occur when the transistor is turned off and has to stand the full rated Vds: ionization from a particle in that condition might start an avalanche process in the channel and lead to the burnout of the transistor. More relevant to ASIC designers is a recent study on lateral diffusion transistors (LDMOS). These transistors are frequently offered in "high-voltage" versions of CMOS technologies, added to the

standard low-voltage transistors for mixed-signal applications, and are frequently used for RF amplifiers or DCDC converters (or more generally in power management ICs). Samples from laboratory-grade LDMOS were observed to burnout during heavy ion irradiations at a Vds below 10 V [12]. Although 3-D simulations showed that this sensitivity was heavily dependent on the specific device construction, investigation on commercial-grade devices should be conducted.

3 The Hardness By Design (HBD) Approach

The idea behind HBD is to remove the radiation tolerance requirements from the hardware level (technology used for the design) and transfer them to the design level. If this allows in principle the usage of any CMOS technology, including state-of-the-art, a specific knowledge of the radiation response of different comparable technologies could help choosing the one requiring less invasive corrective provisions. Figures 2 and 3 indicated how different the radiation response of similar technologies can be. Unfortunately such data are not very common, and their validity in time often doubtful (one should ensure that the natural radiation response of the technology is not altered in time by processing modifications introduced, for instance, to improve yield or lower cost). Different techniques are available to improve radiation tolerance to TID or SEEs.

3.1 HBD for TID Effects

As already reported in Sect. 2.1, the thin gate oxides of modern deep submicron processes are almost free from TID effects. The limitation for the use of these CMOS processes in a radiation environment rather comes from the large density of holes trapped in the thick STI oxide (whose amplitude and effect has been shown to vary widely). Because of this, the typical TID-induced failure mechanism is related to an increase in the leakage current up to a point where basic circuit functionality is lost.

The obvious design-based solution to eliminate any leakage is to avoid the contact between the STI oxide and any p-doped region where leakage current could possibly flow. For the NMOS transistor's edges, where source-drain leakage can flow, this is possible by completely surrounding one of the two n+ diffusions (source or drain) with the thin gate oxide [13]. Several transistor layouts are possible in this respect, some of which are shown in Fig. 5. If ringed layouts (ringed source, ringed inter-digitated) have the advantage of being more compact and allowing for any transistor size, they often require violation of design rules and have been shown to still exhibit some TID effects already in 0.35 μm [14]. In 130 nm implementations, where it is possible to decrease the overlap of the polysilicon over the p-substrate region surrounding the source, radiation-induced leakage current was comparable to the one observed for standard transistors with linear layout (Fig. 6). Despite some

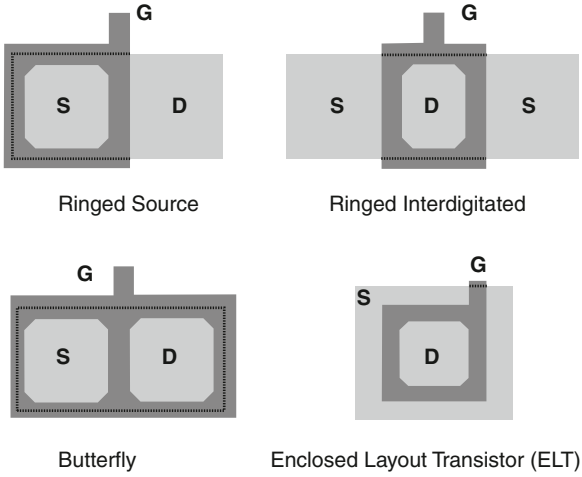


Fig. 5 Transistor layout view for some of the possible NMOS designs eliminating the radiation-induced leakage current between source and drain. The solid line in each design evidences the end of the active area, or the beginning of the STI oxide. The active area under the gate does not get n+ doping, but it is covered by the radiation-tolerant thin gate oxide that surrounds either the source or the drain, or both

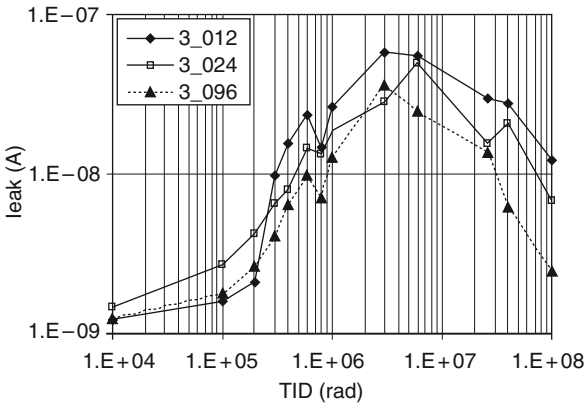


Fig. 6 Source-drain leakage current with TID in NMOS transistors with ringed source layout in a 130 nm technology. The transistors had the same W but varying gate lengths

disadvantages that will be detailed later, the most commonly used layout is the Enclosed Layout Transistor (ELT) where one of the diffusions is entirely surrounded by the other, hence avoiding the violation of any design rule.

Radiation-induced charge trapping in the STI oxide can open leakage current paths between n+ diffusions at different potential (transistor to transistor or n-well to transistor). To prevent this, an effective technique is the introduction of p+ "guardrings", designed as minimum-width p+ diffusions, between the n+ diffusions. These heavily doped p+ regions can not be inverted by trapped holes in the STI. Care must be taken not to draw polysilicon lines above the guardring, because in CMOS processes this would automatically prevent the p+ doping under the polysilicon, hence introducing a weakness. Because of this precaution, the guardring separating the PMOS n-well from the NMOS transistors in logic circuits actually prevents the conventional polysilicon connection between the p and n transistors. To allow for the additional routing (vias to M1), a considerable area is wasted. To

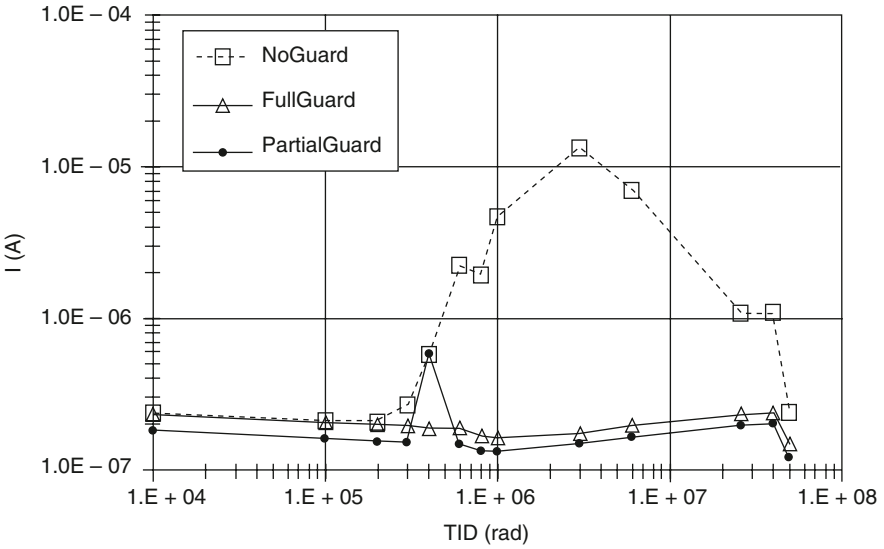


Fig. 7 Vdd to Vss leakage current measured during irradiation on an array of logic cells where the n-well containing the PMOS transistors was not (*NoGuard*) separated from the NMOS transistors, or it was separated by a complete p+ guardring (*FullGuard*) or by a ring broken to allow for direct polysilicon connection between the gate of the N and PMOS transistors (*PartialGuard*)

verify if this is a mandatory penalty to achieve TID tolerance in 130 nm, we have integrated logic blocks laid out in three different ways: with or without full guardring between n-well and NMOS, and with a discontinuous ring broken whenever a polysilicon gate connection was needed. This latter introduced, with respect to the "standard" unguarded layout, only a very small area penalty. The leakage current between Vdd (n-well) and gnd (diffusion of the NMOS) was then measured after irradiation. Figure 7 shows that a partial guard is sufficient in the studied technology.

The combination of systematic use of ELT NMOS transistors and guardrings, often referred to as "radiation tolerant layout techniques", has been demonstrated to be very effective in CMOS processes of different technology nodes [15]. Since it only relies on the natural TID tolerance of modern thin gate oxide, a physical parameter that has virtually no dependence on the particular manufacturing process, it is successfully applicable in all technologies and has no sensitivity on process changes. To date, the largest field application of such HBD approach, hence the best documented, has been in High Energy Physics (HEP) experiments at the CERN Large Hadron Collider (LHC) where a commercial 0.25 μm CMOS technology has been used [16]. Some difficulties and practical considerations from the large experience accumulated at CERN in this domain are summarized in the following.

Using ELT transistor. There are many different possible shapes for an ELT transistor: square, octagonal, square with the corners cut at 45°. Each shape needing to be modeled in a different way [17], it is wiser to use only one shape. Our choice fell on the square shape with the corners cut at 45°, compatible with the ground design rules of most technologies, since it avoids too sharp corners where the electric field

could be excessively large and affect reliability. The details of the chosen shape—for instance the length of the cut in the corners and the size of the inner diffusion—have been decided once and for all transistors, and from that choice we elaborated a formula for the computation of the W/L ratio [18]. The good agreement between the formula and the size extracted from measurements on transistors has been verified for different channel lengths, giving us confidence in the model. Additionally, for a given W/L ratio the ELT transistor has larger gate capacitance than a standard transistor, and correction for that should be introduced at the time of simulating the circuits.

Since there is a direct relationship between the chosen gate length and the minimum gate width of the enclosed transistor, it is not possible to design transistors with aspect ratios smaller than a certain value, as shown in Fig. 8 for a 130 nm technology (the actual minimum size achievable depends on the design rules, namely on the minimum size for a contacted diffusion surrounded by polysilicon). To obtain high W/L values it is sufficient to stretch the base shape in one or two dimensions, without modifying the corners; the calculation of the obtained W/L is straightforward. To have low aspect ratios the only way is to increase the channel length L keeping the minimum size for the inner diffusion: the minimum W/L ratio achievable is about 2.5. Values close to this also imply a considerable waste of area and should be avoided using different circuit topologies.

The evident lack of symmetry in the ELT layout translates in asymmetries in some of the transistor electrical characteristics. In particular, we have observed an asymmetry in the output conductance and in the matching of transistor pairs [16]. Since the gate is annular, the source and drain contacts can be chosen inside and outside the ring of the gate, or the other way round. The measurement of the output

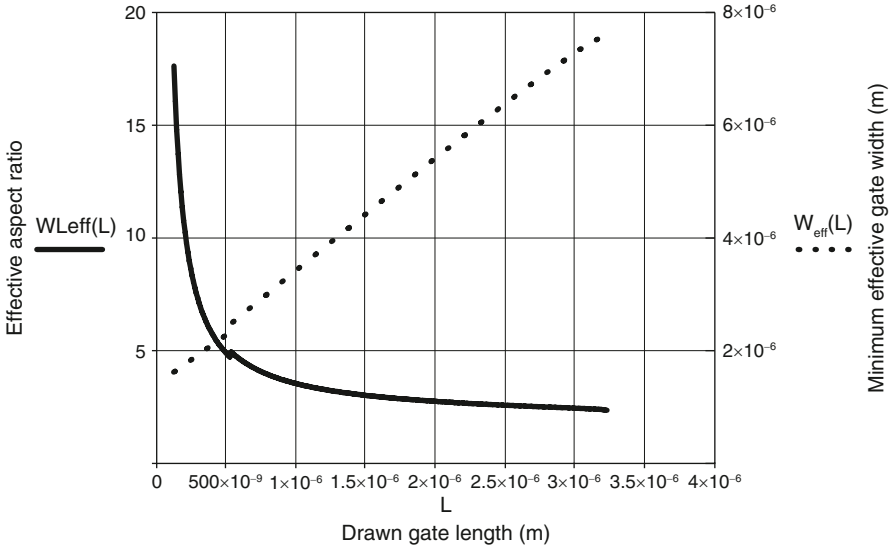


Fig. 8 Minimum effective aspect ratio and gate width for ELT transistors of varying gate length in a 130 nm technology

conductance evidences larger values when the drain is chosen to be inside the ring: the choice of an outside drain lowers the output conductance by 20% (for the shorter gate length) to 75% (for a gate length of 5 μm). In comparison, a transistor with standard linear layout has an output conductance that is close to the mean of the two values measured for the ELT.

Also the matching of ELT transistor pairs reveals the peculiarities of this type of transistor. In the first place, it appears that ELTs have an additional mismatch source, independent of the gate area and never observed before in standard transistors. The mismatch contribution of this additional source depends on the size and shape of the drain electrode, hence the mismatch properties are different for the same transistor when the drain is chosen to be the inner or outer diffusion, another evidence of the lack of symmetry of the transistor.

Area overhead. The systematic use of ELTs and guardrings is evidently an area-consuming practice. As a result, designing with HDB techniques in a given technology implies an important loss of density, in terms of gates per mm^2. In the case of the quarter micron technology extensively used by CERN, the area overhead introduced by HDB was estimated for a number of digital cells. If very simple cells such as inverters impose a 35% penalty only, more complex DFF cells required up to 75% more area. Overall, the penalty for reliable radiation tolerance was estimated to about 70%. The area penalty for analog design, where transistors are rarely minimum size, is typically much smaller.

Yield and reliability. Due to the lack of industrial experience with designs using ELT transistors, a justified concern existed on the possible impact of this layout on the yield and reliability of the ASICs. Yield data can be obtained from the large scale application of HBD in the CERN LHC experiments. In this case, about 100 different ASICs have been designed, some of which produced in quantities above 100,000 pieces. Overall, the production volume for this application exceeded 3,000 8-inches wafers. These ASICs included very different functions: readout for particle detectors, A/D converters, digital and analog memories, system control functions, time-to-digital conversion, transmission of digital and analog data via optical fiber and clock recovery [16]. The measured yield of all these ASICs was comparable to the yield of other standard products run in the same line, demonstrating the absence in yield penalties associated to the use of ELT transistors.

Reliability concerns arise because of the corners used in the ELT transistor design: in those regions the electric field could be more intense and give origin to a larger vulnerability to hot carrier damage. Detailed studies have been carried out to compare the sensitivity of standard and ELT transistors to Channel Hot Carrier (CHC) stress. A first work in the quarter micron node evidenced that the hot carrier lifetime in ELT transistors depends—as it was for the output conductance and the matching—on the choice of the position of the source and drain diffusions [19]. Transistors with the inner diffusion as drain showed about three times lower hot carrier lifetime than standard transistors, which in turn had about three times lower lifetime than ELTs with outer diffusion as drain. On the other hand, a more recent work in 130 nm revealed no significant difference between the two layouts, and also how radiation actually improves CHC resilience in ELT transistors [20].

3.2 HBD for Single Event Effects

Custom SEU-hardened cells. Memory cells and latches used in registers are typically the circuit elements most vulnerable to SEUs, therefore HBD approaches often focus on hardening these cells. A conceptually simple hardening method is to increase the charge needed to upset them (this quantity is commonly called "critical charge"). This can be achieved with the addition of some capacitance to the sensitive nodes, and it is very effective against ionization events triggered by protons and neutrons. For this reason, this technique is used in HEP experiments and for terrestrial applications with high reliability requirements. In an example in HEP, the larger capacitance was integrated either by increasing the size of some transistors, which had also the beneficial side effect of increasing their current drive, or by adding metal-metal capacitors on top of the cells [21]. A reduction of the error rate by a factor of 10 could be easily reached without directly loosing area, but the price to pay was larger power consumption and the loss of two metal layers for routing on top of the cells. An extension to this concept is used by industry to decrease the SRAM sensitivity in terrestrial and avionic applications by more than two orders of magnitude: in that case the additional capacitance is integrated with the addition of a DRAM-like capacitor on top of the SRAM [9].

Another approach relies on the development of a modified cell architecture rendering it less sensitive or—in some cases—insensitive to the charge deposited by the particle strike. Although a large number of designs have been proposed, only a few have actually been used in the field, since additional complexity or associated penalties make most of them unpractical.

Probably the first solution proposed and extensively used to harden SRAM cells is the addition of two large resistors to the cell loop [22]. The cell being formed by two cross-coupled inverters, the function of the resistors inserted between the output of each inverter and the input of the other is to delay the propagation of the signal across the loop. This technique renders the circuit much slower, which is incompatible with the requirements of advanced SRAM circuits. Another cell that has been used in the past is the Whitaker cell [23], which uses the knowledge that particle strikes in n+ diffusions can only induce a change of state when the stored value is a logic 1 (the opposite is true for p+ diffusions). A clever duplication of the nodes where the information is stored creates a cell with four storing nodes, two of which only have n+ diffusions and the other two only p+ diffusions, the two nodes of the same type storing opposite information. In this way, there are always two nodes that are not vulnerable to upset, and a careful connection of the transistors to form the memory loop ensures that in no condition a transient in one of the vulnerable nodes can propagate.

Other dedicated architectures are the HIT (Heavy Ion Tolerant) [24] and DICE (Dual Interlock Cell) [25], both of which have been used in real applications. These cells also use the concept of duplicating the nodes storing the information, and in particular the DICE is very attractive and extensively used because of its property of being compact, simple and hence compatible with the design of high performance circuits in advanced CMOS processes. A schematic of the DICE cell used as a latch is shown in Fig. 9, evidencing how the input of each of the four inverters constitut-

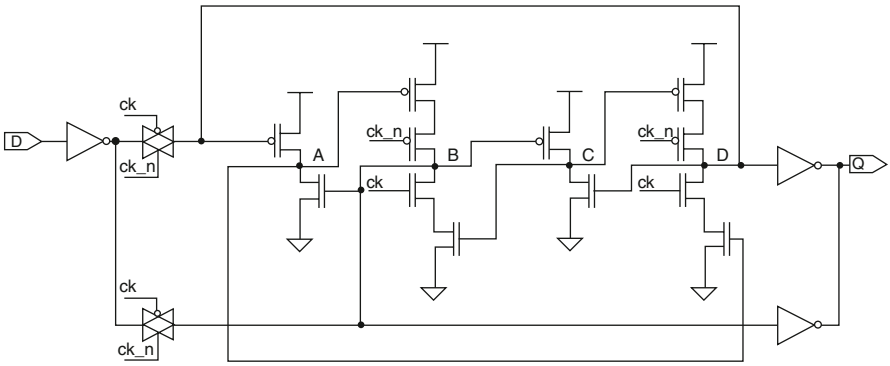


Fig. 9 The DICE cell used as a latch

ing the memory loop is split in two: each of the two transistors of the inverter is connected to a different node. In fact the output of each of the inverters, that are connected as a closed loop, controls the gate of one only transistor (of one polarity) in the following inverter in the loop, and one only transistor (of the other polarity) in the previous inverter in the loop. A particle-induced glitch at the output of the inverter will propagate hence in two directions in the loop, but neither the previous nor the following inverter will in fact change their output and the initial condition will soon be re-established: the error is not latched in the cell. To write in the cell, or to change its condition, it is in fact necessary to write at the same time in two nodes, namely either in the odd or in the even inverters in the loop (nodes B and D in Fig. 9).

The DICE architecture is often used for DFF and register cells, although care must be used in the layout of the cell: the sensitive nodes of the two even and odd inverters have to be spaced sufficiently to avoid a single particle strike to induce a glitch in both, because this would be equivalent to a good writing sequence and would induce the error to be latched. This characteristic is getting a limitation for the use of the DICE cell in 130-90-65 nm CMOS technologies and beyond, where the transistor density is such that all sensitive nodes of the cell are within the reach of a single ionizing particle. Several cells have hence to be "interleaved" in the design, to sufficiently separate sensitive nodes of each cell, but in this case routing resources are wasted in complicated intra-cell connection paths. As an example of the increased vulnerability of the DICE cell, we can compare heavy ion irradiation results of similar designs in 250 and 130 nm [26, 27]. The distance between sensitive nodes was 10 and 2.4 μm respectively in the two technologies, and the work was oriented to the development of radiation-tolerant FPGAs. While the 250 nm design was found practically insensitive to SEU in any tested condition, heavy ion irradiation introduced errors in the 130 nm cell when this was used as a clocked register (no error was instead observed when the cell was used as a memory block, not clocked). Figure 10, illustrates the measured cross-section

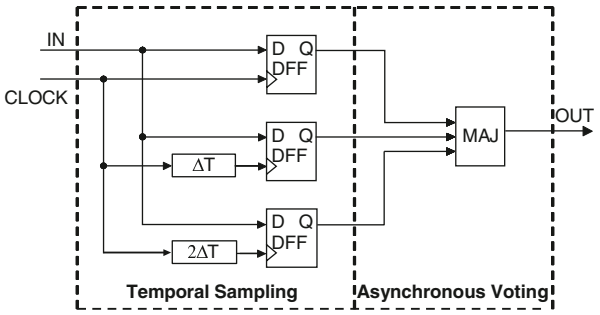


Fig. 10 Simplified architecture of a cell implementing the concept of temporal redundancy

(ratio between the number of errors and the total particle fluence during the test). In particular, at low Linear Energy Transfer (LET) of the ions, errors were only detected when the ion beam was tilted with respect to the sample. In this configuration, the charge deposited by the ion could be collected by the two nodes storing the correct data.

In the 90 nm technology, in a neutron environment, the DICE cell has been measured to give an only tenfold improvement in the error rate with respect to a standard cell. Another inconvenience of the cell is that, although the strike on one node does not latch the wrong data in the cell, the output of the cell can temporarily be wrong during the strike and until the good condition is restored at all nodes. This temporary error at the cell output can propagate to the next cell in some conditions, and eventually be latched somewhere else in the circuit.

A somewhat different approach is based on the concept of temporal sampling [28]. This approach uses the concept of redundancy and builds it inside the basic hardened cell, as shown in Fig. 10. The DFF storing the data is triplicated inside the basic cell, and the output of the three FFs is compared by a voter. In this way, if the content of one of the FFs is upset, the output of the cell is still correct since the voter will output only the value given by the majority of the FFs. The temporal sampling idea has been introduced in the cell to protect it from DSET. If the cell is getting its input from a sequence of combinatorial logic gates, a transient in one of the gates can propagate along the logic chain and reach the input of the cell. A standard DFF would latch the logic value at its input at the moment of the clock transition: if this transition happens in coincidence with the arrival of the transient at the input, the wrong data is latched. To prevent this to happen, the three DFFs in the hardened cell sample the status of the input at three different times, which is implemented by delaying the input by ΔT for one of the FFs and by $2\Delta T$ for another FF. To avoid the transient to be latched, it is sufficient to ensure that ΔT is longer than its duration.

The use of DICE-like cells as registers can be exploited to protect the circuit also from DSET. Since this register requires input data to be written on two nodes, it is possible to duplicate the combinatorial logic such that two independent data paths (with combinatorial logic) are created for redundancy. Each data path actually drives only one input of the SEU-robust register. In turn, each one of the two

data paths is fed by one of the two redundant outputs of the previous register stage. DSET in any of the two data paths will hence not be latched in the register [26, 27].

SEU hardening with redundancy. Another approach to protect the circuit from SEU is to actually add redundancy to the stored information. This can be done in two ways, either by triplicating the cells storing the information (Triple Modular Redundancy, TMR) or by encoding the data and employing Error Detection And Correction (EDAC) techniques.

As already presented for the last cell of the previous sub-section, triplication of the storing cell is a valid way of protecting the content, although very space and power consuming. A single voter circuit can be used to compare the output of the triplicated cell, but in this case the final output could be affected by an error induced in the voter itself. A safer technique is to triplicate the voter as well. In this case, the final output is protected by errors in both the latches and the voters. The voter can either be a separate logic block, in which case its triplication comes at a significant area/power cost, or it can be "embedded" in the latches. Such approach has been used recently in the design of the serializer for a 5 Gbit/s transceiver chipset for optical data transmission in HEP. The speed requirement exceeding the capability of static logic in the target 130 nm CMOS technology, dynamic DFF cells, easily upsettable, had to be used. The latch chain was therefore triplicated, and voters could be embedded in each DFF latch cell as shown in Fig. 11. The final cost for the introduction of every voter was limited to five transistors, and speed specifications were met [29].

Contrary to the TMR technique where each bit of information is triplicated, EDAC techniques require a much smaller information redundancy [30]. EDAC is used very extensively for data transmission and for semiconductor memories, but also whenever data need to be stored reliably (for instance, CD and DVD extensively use it). The information to be stored is encoded by a complex logic block, and some redundant information is added in this process: the larger the number of

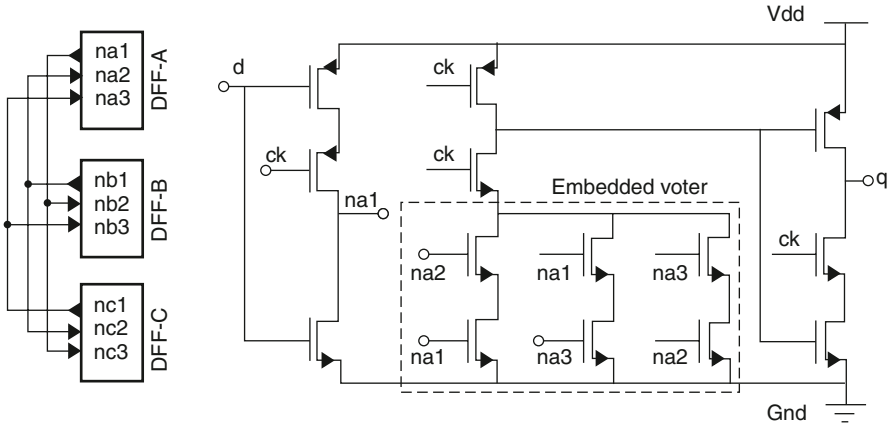


Fig. 11 Schematic of the fast dynamic DFF cell with embedded voter

redundant bits, the safer the original data is from any source of error. At the time of reading the stored information, another complex set of logic operations is needed for decoding. Several codes can be used for EDAC, such as Hamming, Reed-Solomon and BCH, and each of them has different detection and correction capabilities, and different complexity.

Hardening against SEL. In addition to the beneficial effects of retrograde wells and shallow trench isolation typical of modern CMOS technologies, ASIC designers can effectively decrease the circuit sensitivity to SEL with the use of some design practices. A careful examination of the thyristor structure responsible for SEL reveals that it is composed of parasitic resistances and two complementary bipolar transistors. Keeping the resistances small and the gain of the bipolars low effectively reduces SEL sensitivity. This can be achieved for instance by increasing systematically the distance between the two parasitic bipolars, or by using a large number of Vdd and Vss contacts all over the circuit. The latter can be done very effectively by surrounding all regions containing NMOS transistors with p+ guardrings connected to Vss. Therefore, the same layout technique introduced to protect the circuit from TID-induced leakage currents is also very effective to reduce the SEL sensitivity of ASICs [31].

4 Conclusion

Downscaling of CMOS technologies in the 250–65 nm nodes has generally implied a decrease of the sensitivity of ICs to radiation effects. TID effects in the thin gate oxide are practically negligible also for multi-Mrad applications, while leakage currents due to radiation-induced trapping in the STI oxide can sometimes be found to be limited to very acceptable levels (although large variability is unfortunately observed). Sensitivity to both SEL and SEU is also decreasing, but modern fast circuits are prone to digital transients from the combinatorial logic (DSETs). However, due to the all-pervasive diffusion of electronics components in the last decade, and of the large size of memory arrays commonly used, radiation effects have gained the attention of even the large semiconductor manufacturers, with a particular emphasis on SEEs which are now considered relevant also for civilian terrestrial applications.

Whenever circuit reliability is required in the presence of radiation, Hardness By Design techniques leveraging on the intrinsic radiation tolerance of CMOS technologies can be used. These techniques can reliably lead to the production of circuits meeting virtually any radiation specification, still without requiring the use of a dedicated radiation-hard technology. Some of these techniques aimed at protecting circuits against both TID and SEEs have been described, with a large number of examples and references, along with the unavoidable penalties they bring along. These techniques are currently used in the small design communities involved mainly in space, nuclear, avionics and HEP applications, and a limited number of digital libraries systematically using ELTs and guardrings exists. These are typically

limited to a small number of cells (100–200), and are generally not available outside the communities that have produced them. Other than the CERN "radtol" library in 250 nm [16], two libraries in the 180 nm node exist: one from IMEC [32] and one from Mission Research Corporation (MRC) [33]. More recently, radiation results on circuits designed using a library in 90 nm have been presented, evidencing the existence of such a library [34].

As a final remark, it should be noted that new materials are replacing nitrided oxides as gate dielectric materials in very advanced CMOS. The radiation properties of these materials are still unknown. Should they be very sensitive to TID, the limitation in the radiation tolerance of the circuits will again be intrinsic (as it was in the times of large gate oxide thickness) and no HBD technique will enable the overcome of this limitation.

References

1. F. Faccio et al., TID and displacement damage effects in Vertical and Lateral Power MOSFETs for integrated DC-DC converters, in *Proceedings of RADECs 2009*, Bruges (B), Sept 2009
2. N.S. Saks, M.G. Ancona, J.A. Modolo, Generation of interface states by ionizing radiation in very thin MOS oxides. IEEE Trans. Nucl. Sci. **33**, 1185–1190 (1986)
3. V. Re et al., Review of radiation effects leading to noise performance degradation in 100-nm scale microelectronic technologies, in *IEEE NSS Conference Record*, 2008, p. 3086
4. T.R. Oldham et al., Post-Irradiation effects in field-oxide isolation structures. IEEE Trans. Nucl. Sci. **34**(6), 1184–1189 (1987)
5. L. Gonella et al., Total ionizing dose effects in 130-nm commercial CMOS technologies for HEP experiments. Nucl. Instrum. Meth. Phys. Res. Sect. A. **582**(3),750–754 (2007)
6. F. Faccio et al., Total ionizing dose effects in shallow trench isolation oxides, Microelectron. Reliab. **48**, 1000–1007 (2008)
7. P.E. Dodd et al., Production and propagation of single-event transients in high-speed digital logic ICs. IEEE Trans. Nucl. Sci. **51**(6), 3278–3284 (2004)
8. R. Baumann, Single-event effects in advanced CMOS technology, in *Short Course of the NSREC Conference*, Seattle, July 2005
9. P. Roche et al., A commercial 65 nm CMOS technology for space applications: heavy ion, proton and gamma test results and modeling, in *Proceedings of RADECS 2009*, Bruges (B), Sept 2009 (to be published in IEEE Trans. Nucl. Sci.)
10. R. Baumann et al., Boron compounds as a dominant source of alpha particles in semiconductor devices, in *Proceedings of the 33rd Reliability Physics Symposium*, 4–6 April 1995, p. 297–302
11. A.H. Johnston, The influence of VLSI technology evolution on radiation-induced latchup in space systems. IEEE Trans. Nucl. Sci. **43**(2), 505 (1996)
12. P. Dodd et al., Development of a radiation-hardened lateral power MOSFET for POL applications. IEEE Trans. Nucl. Sci. **56**(6), 3456–3462 (2009)
13. D.R. Alexander, Design issues for radiation tolerant microcircuits for space, in *Short Course of the NSREC conference*, July 1996
14. R.N. Nowlin et al., A new total-dose effect in enclosed-geometry transistors. IEEE Trans. Nucl. Sci. **52**(6), 2495–2502 (2005)
15. N. Nowlin et al., A total-dose hardening-by-design approach for high-speed mixed-signal CMOS integrated circuits. Int. J. High Speed Electron. Syst. **14**(2), 367–378 (2004)

16. F. Faccio, Radiation issues in the new generation of high energy physics experiments. Int. J. High Speed Electron. Syst. **14**(2), 379–399 (2004)
17. A. Giraldo et al., Aspect ratio calculation in n-channel MOSFETs with a gate-enclosed layout. Solid-State Electron. **44**, 981 (2000)
18. G. Anelli et al., Radiation tolerant VLSI circuits in standard deep submicron CMOS technologies for the LHC experiments: practical design aspects. IEEE Trans. Nucl. Sci. **46**(6), 1690–1696 (1999)
19. D.C. Mayer et al., Reliability enhancement in high-performance MOSFETs by annular transistor design. IEEE Trans. Nucl. Sci. **51**(6), 3615–3620 (2004)
20. M. Silvestri et al., Degradation induced by X-ray irradiation and channel hot carrier stresses in 130-nm NMOSFETs with enclosed layout. IEEE Trans. Nucl. Sci. **55**(6), 3216 (2008)
21. F. Faccio et al., SEU effects in registers and in a dual-ported static RAM designed in a 0.25 μm CMOS technology for applications in the LHC, in *Proceedings of the 5th Workshop on Electronics for LHC Experiments*, Snowmass, 20–24 Sept 1999, pp. 571–575 (CERN 99–09)
22. J. Canaris, S. Whitaker, Circuit techniques for the radiation environment of Space, in *IEEE Custom Integrated Circuits Conference*, 1995, p. 77
23. M.N. Liu, S. Whitaker, Low power SEU immune CMOS memory circuits, IEEE Trans. Nucl. Sci. **39**(6), 1679–1684 (1992)
24. R. Velazco et al., 2 CMOS Memory cells suitable for the design of SEU-Tolerant VLSI circuits. IEEE Trans. Nucl. Sci. **41**, 2229 (1994)
25. T. Calin, M. Nicolaidis, R. Velazco, Upset hardened memory design for submicron CMOS Technology. IEEE Trans. Nucl. Sci. **43**, 2874 (1996)
26. S. Bonacini et al., An SEU-robust configurable logic block for the implementation of a radiation-tolerant FPGA. IEEE Trans. Nucl. Sci. **53**(6), 3408–3416 (2006)
27. S. Bonacini et al., Development of SEU-robust, radiation-tolerant and industry-compatible programmable logic components. J. Instrum. (JINST) **2**, P09009 (2007)
28. P. Eaton, D. Mavis et al., Single event transient pulsewidth measurements using a variable temporal latch technique. IEEE Trans. Nucl. Sci. **51**(6), 3365 (2004)
29. O. Cobanoglu, P. Moreira, F. Faccio, A radiation tolerant 4.8Gb/s serializer for the giga-bit transceiver, in *the Proceedings of the Topical Workshop on Electronics for Particle Physics (TWEPP) 2009*, Paris, 21–25 September 2009, pp. 570–574 (CERN 2009–006)
30. S. Niranjan, J.F. Frenzel, A comparison of fault-tolerant state machine architectures for space-borne electronics. IEEE Trans. Reliab. **45**(1), 109 (1996)
31. T. Aoki, Dynamics of heavy-ion-induced latchup in CMOS structures. IEEE Trans. El. Dev. **35**(11), 1885 (1988)
32. S. Redant et al., The design against radiation effects (DARE) library, in *RADECS2004 Workshop*, Madrid, Spain, Sept 2004, pp. 22–24
33. D. Mavis, Microcircuits design approaches for radiation environments. Presented at the 1st European Workshop on Radiation Hardened Electronics, Villard de Lans, France, 30th March–1st April 2004
34. D.E. Pettit et al., High speed redundant self correcting circuits for radiation hardened by design logic, in *Proceedings of RADECS 2009*, Bruges (B), Sept 2009

EMC Robust Design for Smart Power High Side Switches

Paolo Del Croce and Bernd Deutschmann

1 Introduction

Integrated circuits often turn out to be the source of interference problems as they have a substantial influence on the electromagnetic compatibility (EMC) of the entire electronic system in which they are used. Because the complexity of EMC problems is increasing with the number of electronic systems assembled (especially in modern cars), the early prediction of the final EMC behavior becomes more and more challenging. To avoid additional costs and time intensive redesign cycles in the IC development, EMC simulations are increasingly used to identify and solve EMC problems in the early design stage. The various electronic systems within a car are composed of several electronic modules which are in most instances directly interconnected by a cable harness. Very often this cable harness acts as receiving and transmitting antenna for RF interferences. Therefore the automotive industry is demanding high standards for the electromagnetic emission and immunity of electronic systems.

Electromagnetic immunity becomes an important issue for the design of ICs especially if the pins of the IC have a direct connection to wires in the cable harness. These wires pick up noise from the electromagnetically polluted environment and transmitting it e.g. by superimposition of the disturbance signal to the power supply to very sensitive functional units of the IC. This often results in undesired behaviors and can cause failures in different ways or even the destruction of the IC. As an integrated circuit has a complex architecture, one single weak point can downgrade the immunity of the entire system. For an effective development, IC designers would need inside information on product behaviour during the test, and this cannot be provided by mean of measurements. In the next chapters, after a short introduction to the IC used for the investigations, an introduction to DPI measurement and simulation technique will be given. Later a design methodology with the aim of designing for a given DPI robustness will be presented. A part of the design

P. D. Croce (✉)
Infineon Technologies AG, Siemensstrasse 2, 9500 Villach, Austria
e-mail: paolo.delcroce@infineon.com

H. Casier et al. (eds.), *Analog Circuit Design,*
DOI 10.1007/978-94-007-0391-9_6, © Springer Science+Business Media B.V. 2011

work with comparison of simulation and measurement results will be also given in the next chapters.

2 High Side Switches in Smart Power Technology

Smart Power is a technology class in which high current, high power MOSFETs are integrated monolithically with a planar IC technology [1, 2]. It is especially suitable for designing switches because high performance with high fabrication yield and low costs can be obtained [3]. Figure 1 shows a commonly used configuration for a high side switch in an electronic control module.

Switches can be classified as high or low side according to their position with respect to the load. Low side switches are connected between the load and GND, while high side switches are connected between the load and the positive supply rail. In this paper only high side switches will be considered. The supply pin of the switch is directly connected to the battery via a cable harness and the output pin is connected by a harness to the load. As mentioned before, the cable harness might act as a receiving antenna for an interference signal that is directly conducted to the power transistor of the switch. Once the interference reaches the switch, it can easily be coupled e.g. via parasitic coupling capacitances into the silicon substrate of the IC, where it can interfere with other components on the same die through parasitic paths [4]. A block level description of a modern high side power switch [5] is shown in Fig. 2. A Serial Peripheral Interface (SPI) is typically embedded for multi-channel switches, while single and double channel switches are generally provided with one digital input per channel.

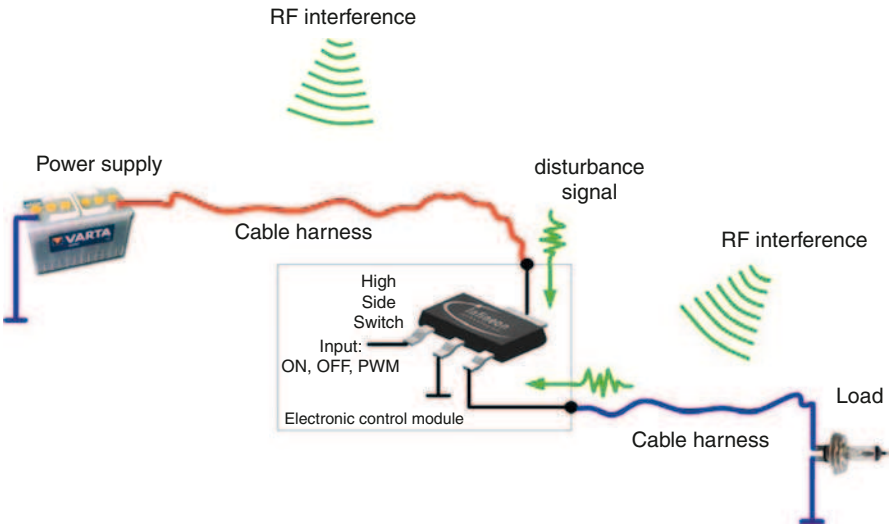


Fig. 1 Typical configuration of a high side switch in automotive application

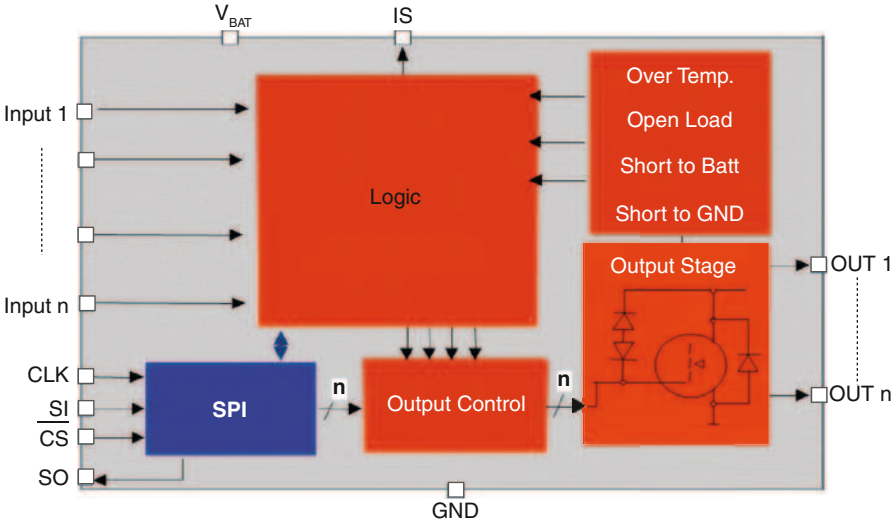


Fig. 2 Typical block diagram of a modern high side switch

3 DPI Test for High Side Switches—Measurement Method

To characterize the electromagnetic immunity of ICs the "IEC 62132: Integrated circuits—Measurement of electromagnetic immunity" standard [6] defines different measurement methods including a direct power injection technique, the DPI method [7]. This technique can be used to inject disturbance signals directly into the pins of an IC. The scope of this measurement method is to characterize the immunity of ICs against conducted RF disturbances. The method guarantees a high degree of repeatability and is therefore often used by semiconductor manufacturers to characterize their products. An example of the DPI setup for a smart power high side switch is given in Fig. 3.

The continuous wave (CW) disturbance signal which is in this case directly injected into the power supply pin is provided by a variable frequency signal generator. The recommended frequency range for the test is 1 MHz to 1 GHz. A power amplifier is further used to amplify the disturbance signal up to a forward power level of 37 dBm which is finally coupled to the supply pin of the IC under test. For this purpose a coupling capacitor of 6.8 nF (which is also used to isolate the DC supply voltage of the IC from the output of the amplifier) is used. In order to prevent the disturbance signal from being diverted via the power supply or the load to GND several decoupling networks (consisting of a 5 uH inductance) are used. These inductances also represent more or less the typical impedance of the loops provided by the cable harness in the vehicle. Together with the 150 Ω impedance network, which represents the typical antenna impedance of a cable in the har-

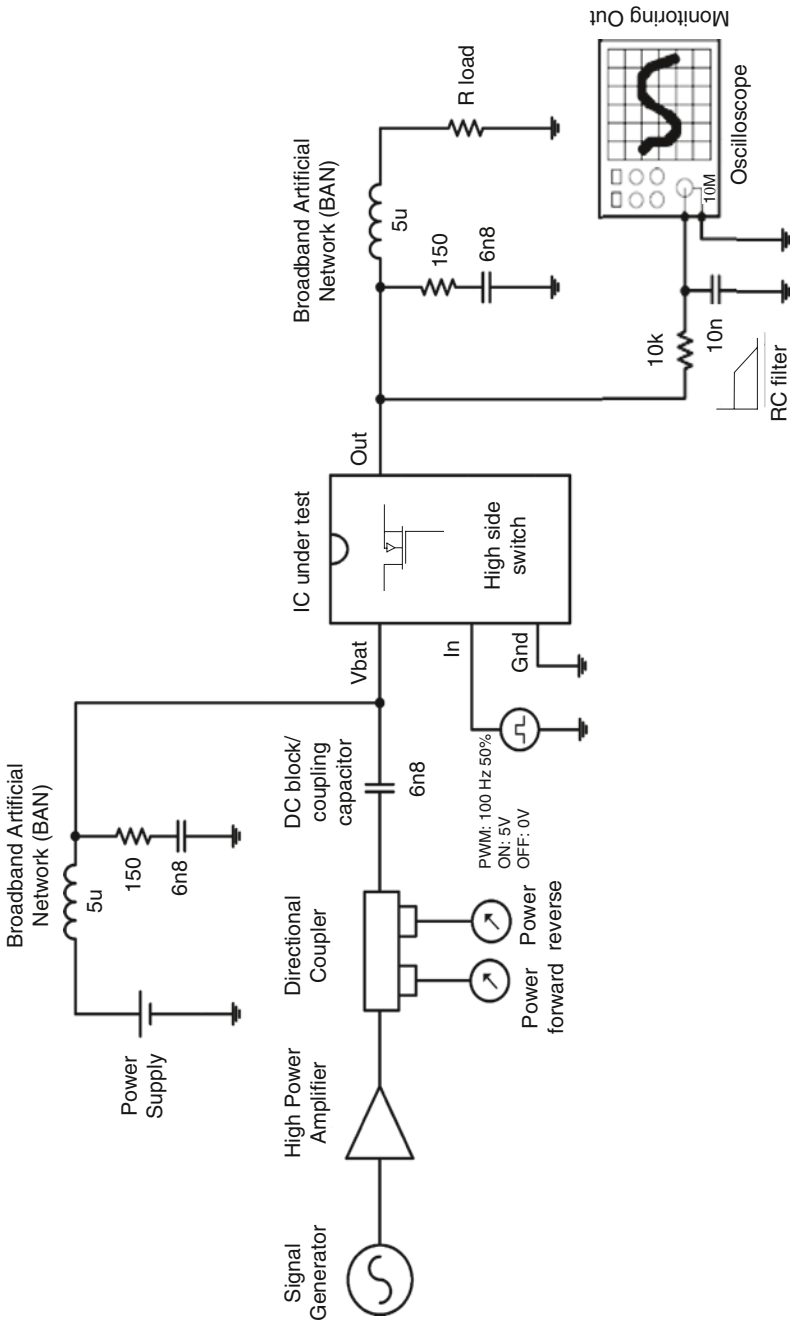


Fig. 3 DPI measurement/simulation setup

ness, they form the broadband artificial network (BAN). For the measurement a directional coupler was used to measure the forward and reflected power by means of two power meters. In our case only forward power is considered and is used as an indicator for the immunity level to which the IC under test functions within the predefined conditions. However, only a small amount of the forward power is dissipated by the IC under test, the remaining being reflected and dissipated in other discrete components.

3.1 Operation Modes of the IC Under Test

Generally the immunity test for switch products is performed with the IC operated in three different conditions: OFF, ON, PWM (pulse width modulation). In OFF mode the switch is open and the load is permanently not supplied; during DPI the device must remain stable, maintaining the load isolated from supply. In ON mode the switch is closed and the load permanently supplied; during DPI the device must remain stable, providing supply to the load. In PWM mode the switch drives the load sequentially OFF-ON-OFF with a frequency of 100 Hz and a duty cycle of 50%. During all DPI tests the device must provide a proper output status with the right timing to the load and keep all signals within predefined tolerance ranges. Robustness in each operation mode requires a careful analysis and specific circuit solutions. In this paper the optimization of the robustness in the OFF mode is considered and will be discussed further.

4 DPI Test for High Side Switches—Simulation Approach

In order to pre-define the final EMC performance of an IC, immunity simulations are often used to identify and solve interference problems in the early design stage. Based on the described IC level measurement technique several attempts to predict the immunity behavior of an IC by performing DPI simulations were conducted in the past. In [8], for example, a complete simulation model of a DPI setup which includes the whole measurement setup as well as the IC and its direct environment is shown. In this simulation model conductive and dielectric losses as well as the directional coupler, the injection probe, the injection capacitor and the power supply are considered. It was shown that an accurate prediction of the immunity levels of an IC under test is possible, if all the losses and external influence are taken into consideration. For the investigations in this paper, a simpler approach was used. The simulation test-bench only considers forward power provided to the particular IC pin, as shown in Fig. 4.

The disturbance signal source, amplifier, external coupling and decoupling elements with relevant parasitic were modeled. The 6.8 nF coupling capacitor e.g. was modeled by a series resistor-inductor-capacitor (RLC) network. For the BAN a 5 uH inductance in combination with a 150 Ω resistor in series to 6.8 nF capacitor

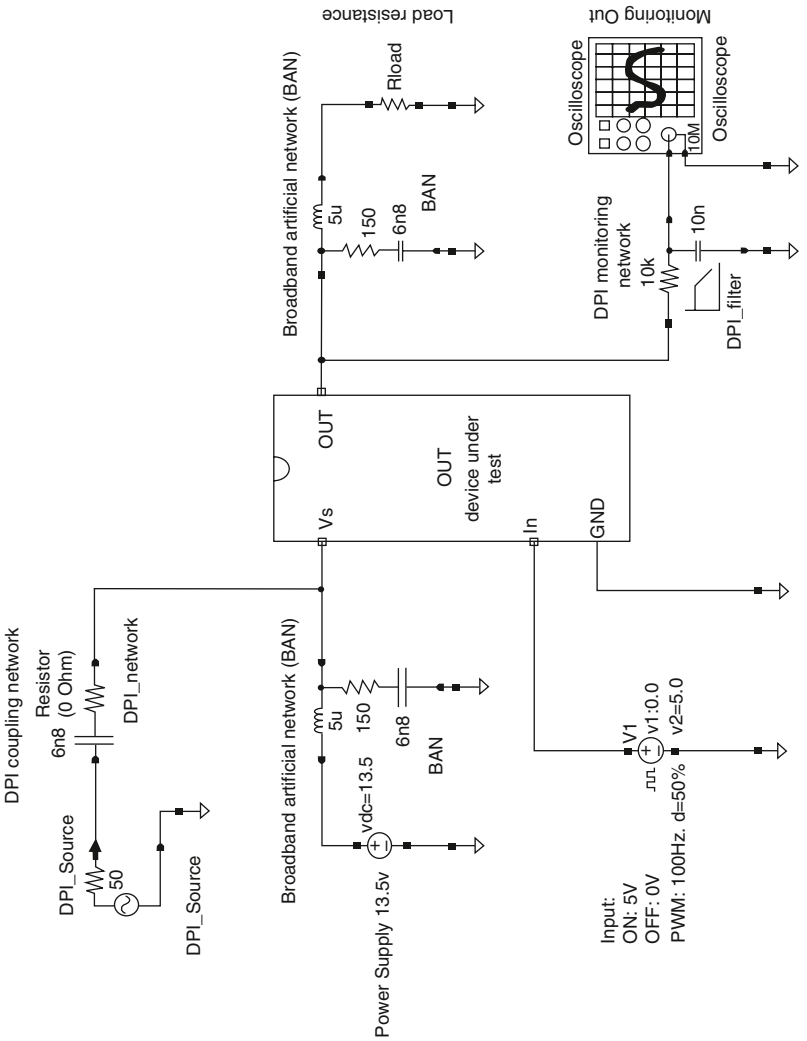


Fig. 4 DPI simulation/verification test-bench for 1-channel high side switches

as a model of the antenna impedance was used. In this simulation test-bench, the output signal can be monitored by the simulation model of an oscilloscope. To filter the disturbance signal from the output an RC filter (10 kΩ/10 nF) has been used. All the relevant signals e.g. the voltage at the battery supply or the load current can be directly probed at the corresponding nodes. Instead of a full chip level simulation where all circuit elements are included, a simplified test-bench, which only contains the most relevant parts, was chosen to keep the simulation time to a minimum. Selected blocks are the power switch, switch driver and the impedance reduction circuitry that will be explained in more detail in the next chapters.

5 DPI Robustness in Off State—Problem Description

Referring to Fig. 1 when the digital input signal has the logic value zero, the high side switch is supposed to be open, isolating the load from battery so that load voltage and current are also zero. Going more in detail, the power transistor shown in the output stage of Fig. 2 must be kept off as long as the input stays at zero. This functionality is coded into a logic and actuated by means of an output control, which provides $V_{GS}=0$ V to the power transistor in the output stage. Leaving aside design related problems and the circuit complexity, this, as any other function can be provided only for battery voltages above a given minimum value $V_{BAT(MIN)}$. Below $V_{BAT(MIN)}$, $V_{GS}=0$ V cannot be ensured anymore by design and the power transistor is in an undefined state. In such a floating condition isolation of the load from battery cannot be guaranteed anymore and a DPI test fail is likely to happen. Such a low voltage condition has to be expected during DPI test because impedance mismatches lead to large voltage swings of the disturbance signal. The signal generator, directional coupler and coupling path have about 50 Ω impedance, while the supply pin of high side switches in off state (chip in stand-by, load isolated from supply) has much higher impedance to GND. In Fig. 5 the superimposing of a sinusoidal

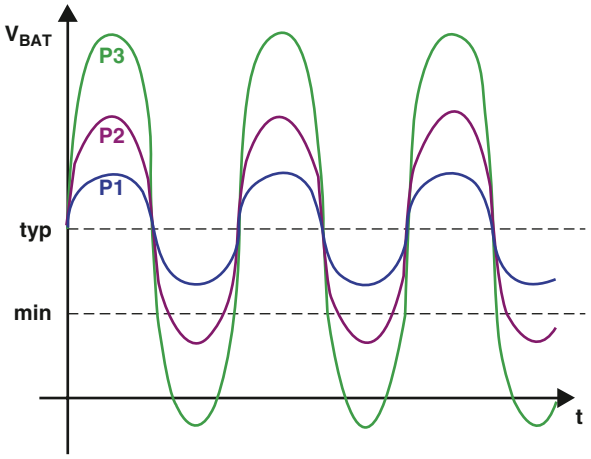


Fig. 5 Superposition of disturbance signal to the typical supply voltage during DPI in off state; increasing injected power from P1 to P3

disturbance signal with the power supply voltage at the supply pin is shown. The typical DC value of the supply (in our case 13.5 V) is shifted up and down with the frequency of the disturbance signal. For higher injected power levels the battery voltage can even go below GND, generating substrate currents that further endanger the functionality of the chip. Additionally, the very high voltage swing can couple AC noise through parasitic capacitances into circuit blocks which might change their state and induce failures.

The scenario is very difficult to simulate due to its full-chip and layout-dependent nature. Further complexity is added from the off-chip environment, which additionally influences the impedance on V_{BAT} line. In conclusion, the main issue in DPI testing in the off state is the RF voltage at the chip supply because of high and uncertain impedance at V_{BAT} pin. In the following pages a method for overcoming these issues and setting up reliable simulations will be discussed.

6 Design Methodology—Pin Impedance Design

Defining V_{BAT} line impedance in an early project state has a relevant impact on chip design and simulation strategy in respect to DPI robustness. It allows estimating IC external and internal supply noise by mean of simulations. When the internal supply noise is known, each sub-block can be designed and tested independently for DPI robustness. Through this approach complex full chip simulations will be used as final cross-check and not as main circuit debugging tool. For preventing issues associated with low battery voltage or substrate current, impedance must be trimmed down in the initial phase of the IC design as long as at target injected power level the battery ripple will look like case P1 in Fig. 5. Pin impedance reduction at chip level cause also off chip impedances to be unimportant and give the designer a free parameter for limiting the maximum RF voltage at the given target DPI immunity. The topology of the circuit for reducing V_{BAT} pin impedance during DPI that is proposed is shown in Fig. 6.

The very simple circuit architecture makes use of three principle elements: M1, D1 and I1. A power transistor M1 with drain and source connected respectively to V_{BAT} and GND set the pin to pin impedance. Its gate is driven from a non-linear current source I1 with rectifying diode D1 in parallel for controlling the pin to pin impedance value. Both capacitances C_{GD} and C_{GS} are MOS parasitic, but have an active role in the control scheme. The working principle is as follows: C_{GD} and C_{GS} work as capacitive voltage divider turning M1 on by means of the rectifying effect of D1 during RF power injection. The non-linear current source works against C_{GD}, avoiding M1 turning on at battery spikes or during battery voltage ramp up. This further ensures M1 is in off state in absence of DPI. Properly sizing I1, the battery voltage ripple that triggers the switching on of M1 can be defined. The verification of this impedance reduction circuit was carried out based on the described simulation test-bench shown in Fig. 4, using the impedance reduction circuit instead of the high side switch IC. Assuming a nominal battery voltage of 13.5 V, M1 switch on

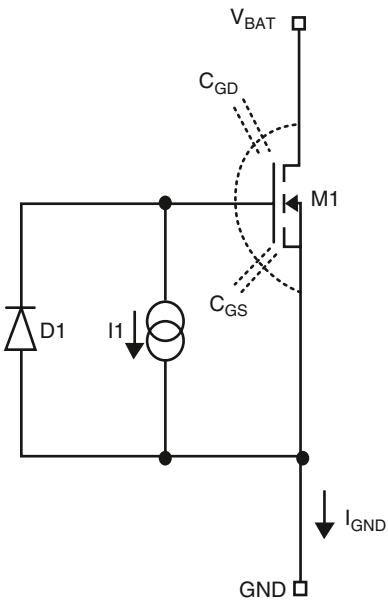


Fig. 6 Proposed circuit topology for pin to pin impedance reduction

threshold has been adjusted to ensure $V_{BAT} > 0$ at any time in the lower frequency range for DPI up to a value of 37 dBm. In Fig. 7 simulation results of the impedance reduction circuit are shown for a frequency of 1.5 MHz and a disturbance signal of 37 dBm. Battery voltage and current flowing via M1 to ground are visible. Figure 8 shows the result of the same simulation at 10 MHz. In both cases the overall picture

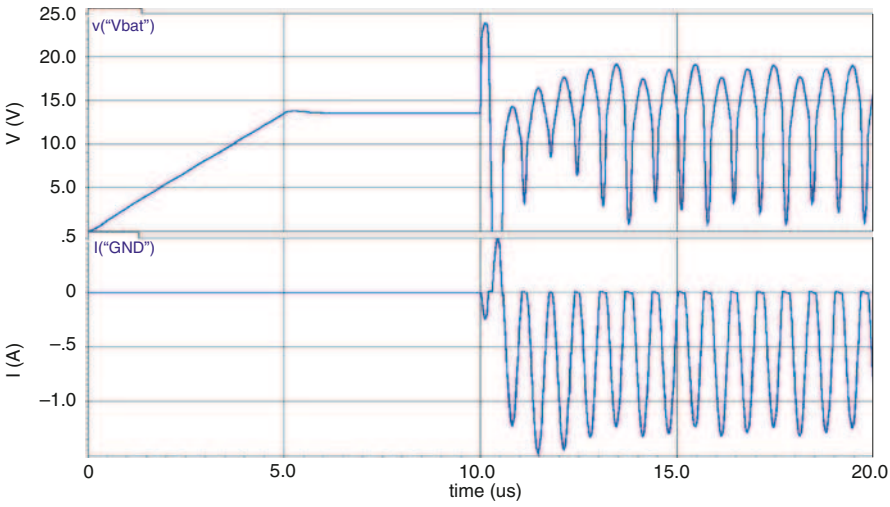


Fig. 7 DPI simulation result of the impedance reduction circuit for 1.5 MHz and 37 dBm (V_{BAT}, I_{GND})

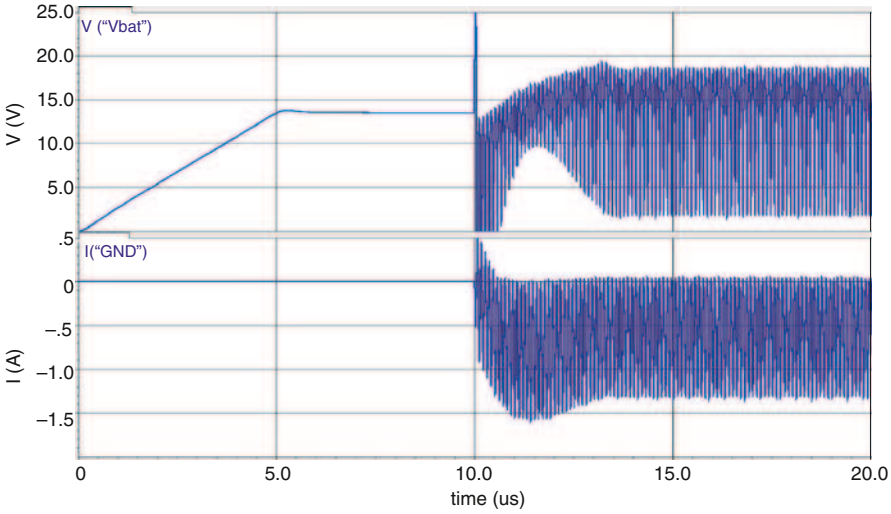


Fig. 8 DPI simulation result of the impedance reduction circuit for 10 MHz and 37 dBm (V_{BAT}, I_{GND})

remains nearly unchanged. During the first 5 µs battery voltage is ramped-up to the nominal value of 13.5 V and with 10 µs delay DPI starts. Maximum voltages are kept below 20 V while minimum values are above 1 V and the average voltage is kept close to its nominal value. According to the simulation, the circuit is capable of limiting pin impedance during DPI, with M1 driving the disturbance signal from the battery pin to GND.

The impedance over the frequency range 1.5–700 MHz has been simulated and results are presented in Fig. 9. The circuit offer quite low AC impedance, decreasing with frequency increase. Because model accuracy in the gigahertz range is reduced, the trend over frequency presents a possible danger. M1 could lock in a low ohmic condition driving direct current (DC) from the supply line to GND, inducing overheating and device destruction.

The problem could be solved designing a frequency dependent current source I1, which compensates the natural behaviour of the proposed circuit. Nevertheless for simplicity no impedance compensation has been introduced. The methodology will be validated in the range 1–10 MHz, and extended to the entire frequency range later on. With the power switch remaining in the off condition (chip is in stand-by and loads are isolated from the battery) M1 will divert nearly all the disturbance energy that enters the chip, consequently its size must be taken into consideration to be able to handle the power dissipation. Also the connection of M1 to V_{BAT} and GND pins needs particular care because of the high current flowing. In the high frequency range, parasitic capacitances at chip level will additionally decrease the supply pin impedance to GND reducing power dissipation on M1. The impact on chip area of the impedance reduction circuit is in the range of fractions of square millimeter, but it is not that relevant since the device M1 can also be used as electrostatic discharge

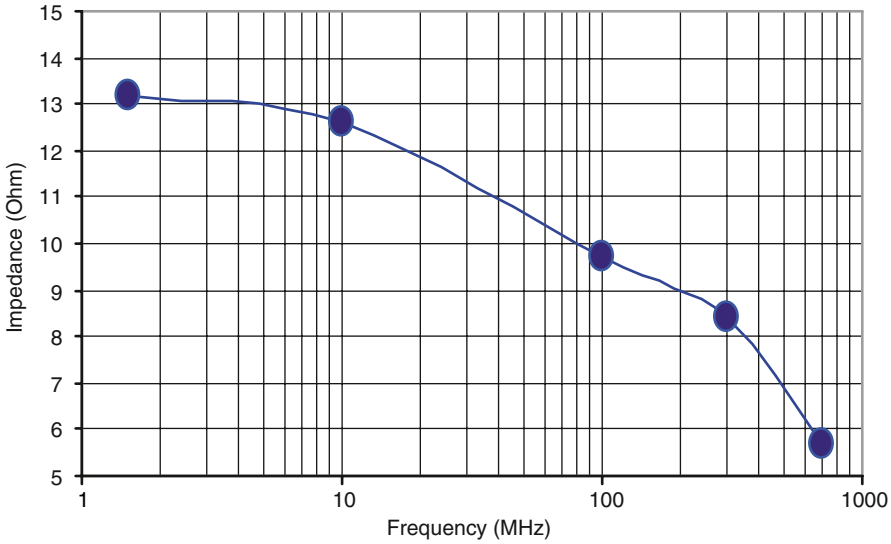


Fig. 9 Simulated V_{BAT}—GND impedance of the impedance reduction block

(ESD) protection as well as over voltage protection, as it was done in this project. By sizing the impedance reduction circuit, the ripple of the disturbance signal at V_{BAT} pin can be optimized. Knowing the ripple at chip supply, internal supply noise can be derived and circuit blocks within the chip can be simulated and enhanced for DPI robustness. E.g. considering a block specific $V_{BAT(MIN)}$ in the range of 5 V, a tank capacitor could be implemented to get over the ~100 ns where $V_{BAT} < 5$ V (see Fig. 7). At the end of the development phase single improvements can be verified together running a chip level simulation which includes all the relevant circuits blocks.

Fig. 10 shows the chip level simulation test-bench used, where all circuit blocks that have a significant influence on DPI robustness of the high side switch are considered. Results from chip level simulations are shown in Fig. 11 and in Fig. 12.

According to application requirements the following external components have been used: load resistor $R_{LOAD} = 3.3\ \Omega$, sensing resistance $R_{IS} = 3.3\ K\Omega$, ground resistance $R_{GND} = 0\ \Omega$. Coupling, decoupling and BAN are as previously specified. Comparing Fig. 11 with Fig. 7 it can be seen that high frequency ripple at V_{BAT} is further reduced because of increased overall parasitic capacitance at this node which is created by additional circuit blocks. Still most of the injected power is dissipated from the impedance control transistor M1. For checking switch isolation capability at the output a monitoring voltage "Vout_monitor" has been defined. Figure 12 shows simulated voltage values of "Vout_monitor" for the high side switch in OFF mode. The lower and the upper curves correspond to the behaviour with and without pin impedance control respectively. If this block is missing the disturbance will find its way to GND through the output stage interfering with the load and degrading switch isolation performance.

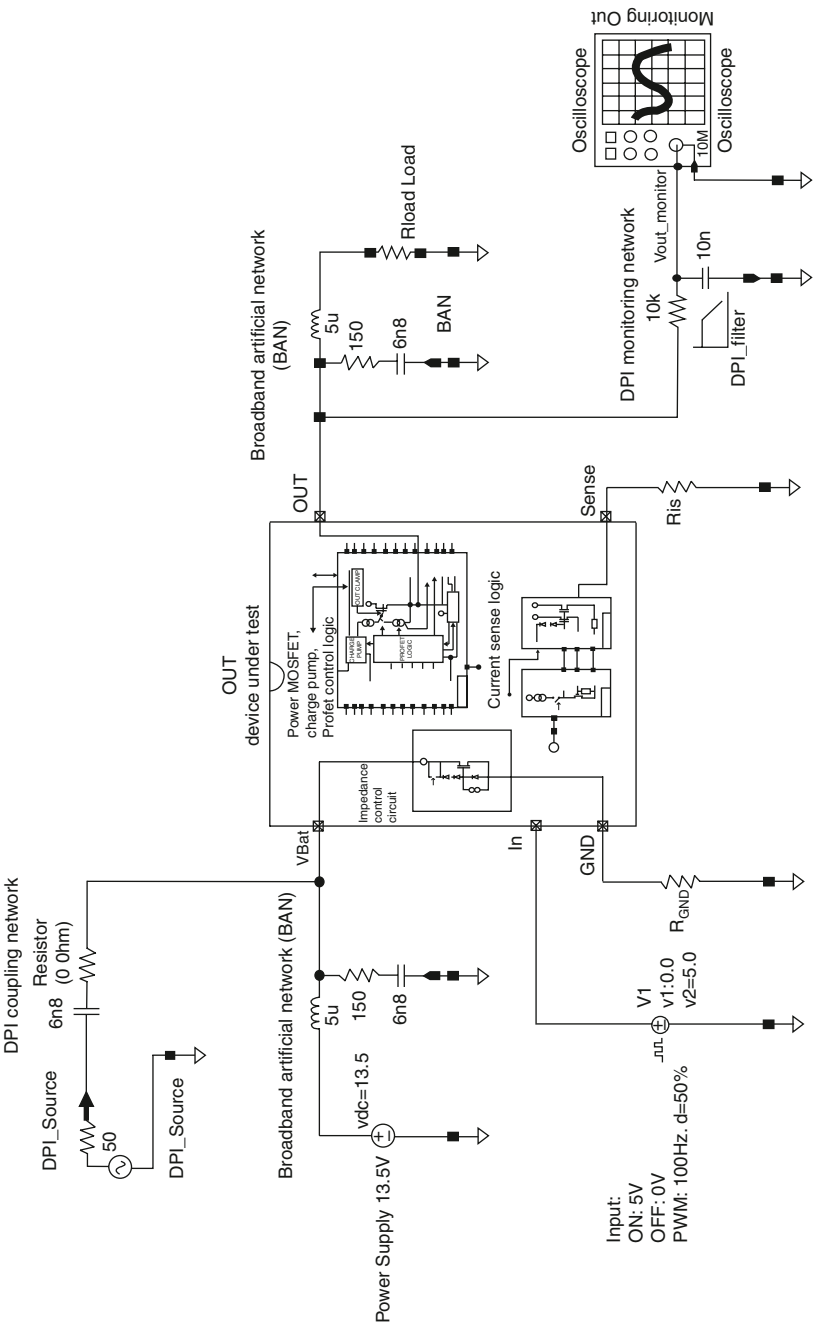


Fig. 10 Chip level verification test-bench including all relevant circuit blocks

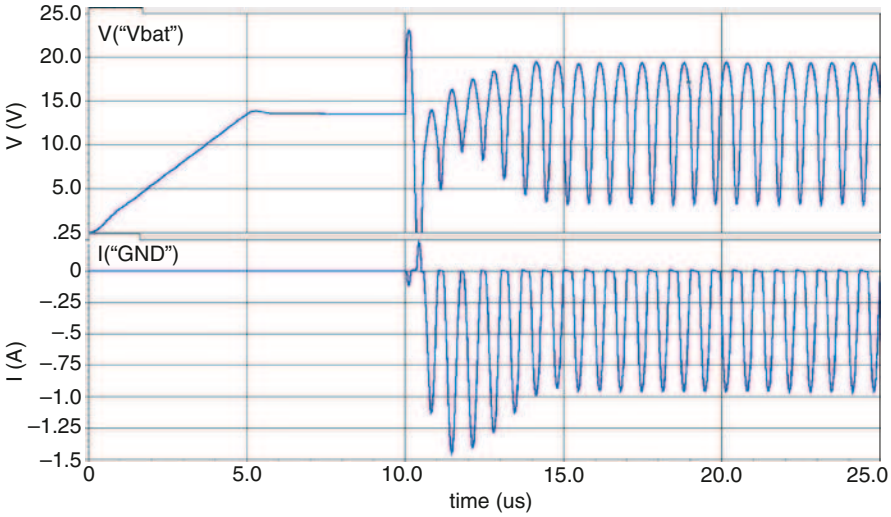
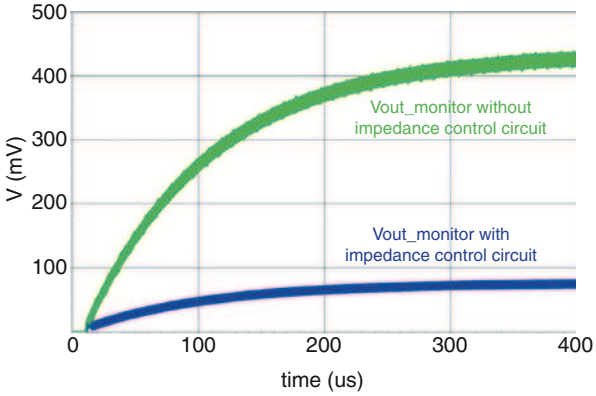


Fig. 11 DPI simulation result of the chip including all relevant circuit blocks for 1.5 MHz and 37 dBm (V_{BAT}, I_{GND})

Fig. 12 DPI simulation result of the output voltage "Vout_monitor" for 1.5 MHz and 37 dBm, with and without pin impedance control circuitry



7 Measurement Results

In this chapter DPI characterization results of the high side switch are shown. The oscilloscope picture in Fig. 13 shows the superimposition of 37 dBm disturbance signal to the nominal supply voltage of 13.5 V at V_{BAT} pin. Comparing the measurement with the simulation result in Fig. 11 we can see a good match especially for signal waveform and min.-max. amplitude of the disturbance signal.

Due to the implemented impedance reduction circuitry $V_{BAT(MIN)}$ is successfully kept to values higher than 3.5 V which helps in keeping the power transistor in a defined state. As a proof of the switch isolation capability during DPI, "Vout_monitor" has also been checked. The signal is shown in Fig. 14. Its mean value is about 65 mV

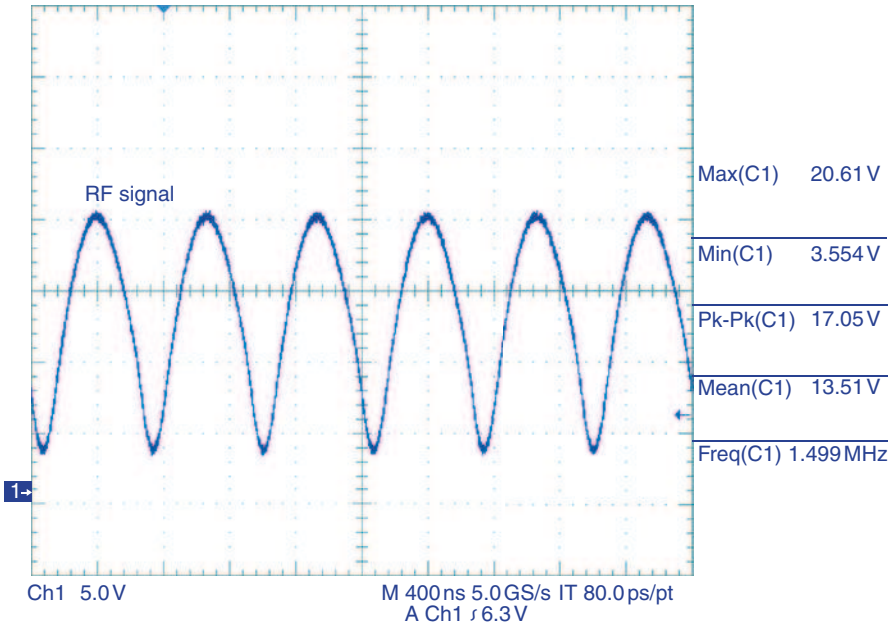


Fig. 13 Disturbance signal applied at V_{BAT} pin; 1.5 MHz, 37 dBm DPI

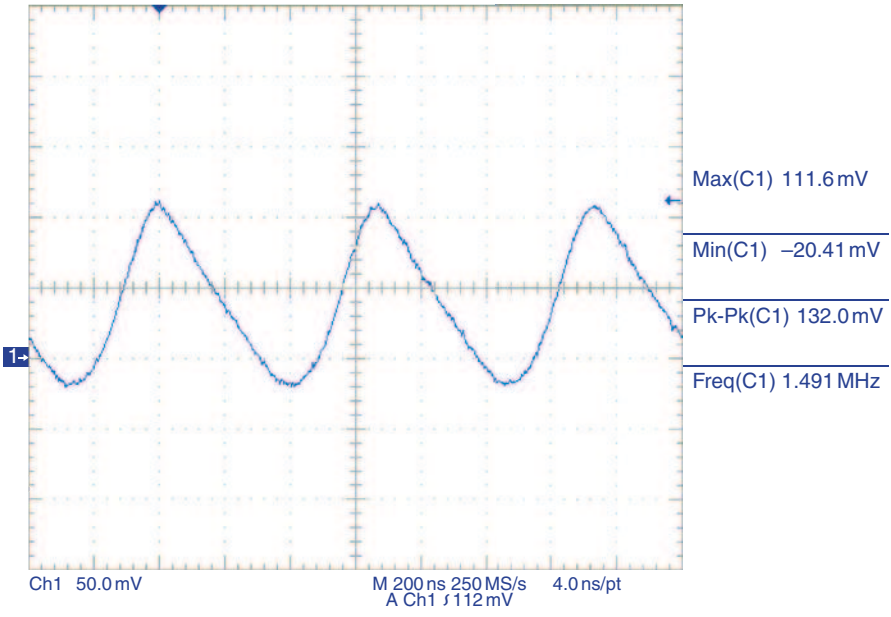


Fig. 14 "Vout_monitor" signal measured during 1.5 MHz, 37 dBm DPI

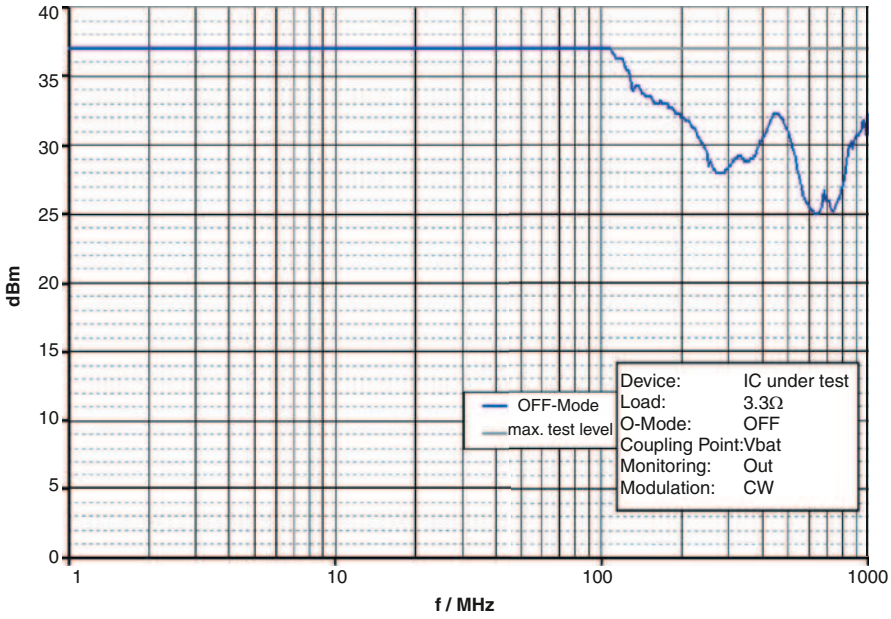


Fig. 15 Susceptibility characteristic of the IC under test for DPI into V_{BAT} up to 37 dBm

while 75 mV were expected (see Fig. 12). This proves the accuracy of simulated results at 1.5 MHz. A comparable level of precision is given at 10 MHz. The whole susceptibility characteristic of the chip for the output in the off state can be found in Fig. 15. In this case the disturbance signal was swept from 1 MHz to 1 GHz. At each frequency step, the amplitude of the disturbance signal was increased either to the maximum test level of 37 dBm or to the level where the output signal went out of the pre-defined tolerance range of ±1.35 V (10% of its nominal value). As can be seen, the simulated robustness in the low frequency range can be confirmed by measurement. For frequencies higher than 100 MHz we expects no degradation of the performance of the impedance control circuit as shown in Fig. 9, but as can be seen in the diagram the susceptibility characteristic is decaying down to 25 dBm. The malfunction is most probably caused by failure modes which are not related to the voltage ripple at V_{BAT} pin and can be ascribed either to the extremely low pin impedance or to failure modes not modeled in this work.

8 Conclusion

For this work the DPI measurement test bench of a smart power high side switch product has been transferred to the simulation environment, as pre requisite for IC robustness verification and optimization. An optimization methodology based on pin impedance control has been proposed and applied during product development.

Simulation and measurement results have been compared to validate the approach. Effectiveness of the methodology in the low frequency spectrum can be confirmed, while predicting high frequency behavior will require further modeling effort. In particular, parasitic inductances have to be evaluated to account for resonances. Device to device silicon parasitic has to be considered to account for substrate current induced fails. Device simulation accuracy has to be guaranteed over the entire DPI frequency range. Enhancement of IC modeling is out of the scope of the paper, but is an essential step for successfully extending the proposed design methodology to the entire DPI test frequency spectrum.

References

1. R.S. Wrathall, D. Tam, Integrated circuits for the control of high power, in *Proceedings of IEEE International Electron Device Meeting IEDM,* 1983, pp. 408–411
2. J. Einzinger, L. Leipold, J. Tihanyi, R. Weber, Monolithic IC power switch in automotive applications, in *Proc. IEEE Int. Solid State Circuits Conference,* 1985, pp. 22, 23
3. A. Elmoznine, J. Buxo, M. Bafluer, P. Rossel, The smart power high-side switch: description of a specific technology, its basic devices, and monitoring circuitries. IEEE Trans. Electron. Devices. **37**, 1154–1161 (1990)
4. F. Fiori, Prediction of RF interference effects in smart power integrated circuits, in *Proc. IEEE International Symposium on Electromagnetic Compatibility,* 2000, vol. 1, pp. 345–347
5. A. Graf, Smart power switches for automobile and industrial applications, in *Proc. VDE ETG Conf, Contact performance and switching,* 2001, pp. 1–8
6. International Electrotechnical Commission, *IEC 62132-1, Integrated circuits—Measurement of electromagnetic immunity, 150 kHz to 1 GHz—Part 1: General and definitions,* 1st edn. (IEC, Geneva, 2006)
7. International Electrotechnical Commission, *IEC 62132-4, Integrated circuits—Measurement of electromagnetic immunity 150 kHz to 1 GHz—Part 4: Direct RF power injection method,* 1st edn. (IEC, Geneva, 2006)
8. A. Alaeldine, R. Perdriau, M. Ramdani, J.-L. Levant, M. Drissi, A direct power injection model for immunity prediction in integrated circuits. IEEE Trans. EMC. **50**(1), 52–62 (2008)

Part II
Sigma Delta Converters

In this part new advances in sigma delta AD converters are discussed. Both sampled-data as continuous time topologies are presented, but also some new trends such as VCO based and comparator based topologies will be analyzed.

The first two contributions are sampled-data topologies. The first focuses on the use of noise coupling, based on the correlation effects of the quantization noise in the loop. As such it is proven that the overall signal-to-noise ratio can be improved, and that those techniques can be combined with time-interleaved techniques to further improve the noise shaping and bandwidth of the converter.

The second contribution is concentrated on techniques to reduce the oversampling ratio. As such sigma delta topologies can become closer to Nyquist-rate A/D converters and as a result they can achieve a higher signal bandwidth. In this work an 8th order Mash topology with only an over sampling ratio of 3 is demonstrated.

The next two contributions are sigma delta topologies based on new trends. The first one is a comparator based sigma delta. Instead of using power hungry OTA's, it is proposed to use comparators. Of course this results in some challenges in the used charge-pumps, but an implemented 90 nm CMOS design proves that the technique can achieve state-of-the-art low voltage performances.

The next work demonstrates a trends towards the use of VCO based topologies. Instead of a multi-bit quantizer, a VCO is used as quantizer. The information however is now in the time domain, and as such the critical building block becomes now the time-to-digital circuit. In this work this techniques is demonstrated in combination with a continuous time sigma-delta topology and resulted in a 130 nm CMOS design.

Besides sampled-data also continuous-time sigma delta topologies are discussed towards wideband performances. In this contribution a multi-bit continuous time sigma-delta is analyzed and discussed. An achieved 25 MHz bandwidth in combination with a SNDR of 67 dB results in one of the highest bandwidth to-date sigma delta topologies.

The last contribution is about DA converters. Also in this area sigma delta techniques are popular. In this contribution, challenges concerning high performance

audio DA converters are discussed. Especially in combination with the required smoothing filter and jitter effects due to the conversion towards continuous signals are analyzed. Examples in combination with hybrid SC continuous-time reconstruction filter topologies are presented.

Michiel Steyaert

Noise-Coupled Delta-Sigma ADCS

Kyehyung Lee and Gabor C. Temes

1 Introduction

In wired and wireless communication, digital video and other consumer electronics applications, there is increasing demand for wideband and high resolution data converters with good power efficiency. ΔΣ ADCs can deliver high resolution and linearity with lower power consumption in the MHz-range signal band [1–4]. It is hence the ADC architecture of choice in many applications. A ΔΣ ADC provides high resolution and linearity while using only a low-resolution quantizer by taking advantage of oversampling and noise shaping. There are three key design parameters: (1) oversampling ratio (OSR); (2) loop filter order L; (3) quantizer resolution M. Increasing any of these improves the resolution under ideal conditions. However, there are limitations associated with each design parameter. In a wideband modulator, the OSR is limited by the sampling frequency for a given signal bandwidth. Higher sampling frequency in a modulator implies reduced settling time for its components, and its amplifiers need more power to keep the same settling accuracy. To meet demanding design targets with low OSR, high loop filter order and quantizer resolution are required. The loop filter order determines the effectiveness of its noise shaping and is limited by the modulator stability in a single-loop modulator. The quantizer resolution is typically limited up to 4 to 5 bits due to the exponential growth in the hardware complexity and the power dissipation of the flash ADC needed. Using a tracking ADC, VCO-based ADC, or pipelined ADC as a quantizer may address this issue [5–7].

There are two possible realizations of a ΔΣ modulator. Discrete-time (DT) realization, based on switched capacitor (SC) circuitry, is the traditional one, and continuous-time (CT) realization is getting popular due to its superior power efficiency and inherent antialiasing filtering for wideband modulators. In a DT modulator, clock jitter is not an issue if the time constants of the capacitors and switches are sufficiently small. In CT modulators, however, clock jitter is unavoidably converted

K. Lee (✉)
Conexant Systems, Newport Beach, CA 92660, USA

H. Casier et al. (eds.), *Analog Circuit Design,*
DOI 10.1007/978-94-007-0391-9_7, © Springer Science+Business Media B.V. 2011

into noise, and thus represents a major limitation on the SNR performance [8]. On the other hand, due to the settling requirement in the stages of a DT modulator, its opamps need to have a larger unity-gain bandwidth (GBW) than those of a CT one. Hence, CT ADCs can provide wider signal bands but lower SNDR than DT ones with comparable complexity and power dissipation [9–11]. The loop filter coefficients in a DT modulator are accurately defined by capacitor ratios, while those of CT modulator experience a large variation with process, voltage, and temperature and tuning circuitry is typically required. In summary, compared with CT modulators, a DT modulator offers robust operation under process variation, and is less sensitive to clock jitter. Its accuracy relies on precise capacitor matching, which is relatively easy to achieve. The disadvantages of DT modulator are higher power dissipation, and a need for an antialiasing filter. In contrast, CT modulators usually do not require any antialiasing filtering, since this is inherent in the continuous-time loop filter [12].

In the design of the devices discussed in this chapter, DT realization was chosen to avoid the effects of element mismatches and clock jitter. A low-distortion modulator topology is well known for its reduced signal swing inside the loop filter, which allows superior modulator linearity and relaxed opamp design [13, 14]. It will be shown that both the power efficiency and linearity of the low-distortion modulator can be improved using quantization noise coupling and time interleaving. Hence, in the design of the proposed ΔΣ modulator, we relied on architectural innovations to achieve good performance combined with low power dissipation.

This chapter is organized as follows. Architectural improvements for the conventional low-distortion ΔΣ modulator are proposed in Sect. 2. A detailed description of the design of prototype ΔΣ ADCs follows in Sect. 3. Section 4 presents the measured results. Finally, conclusions are drawn in Sect. 5.

2 Noise-coupled ΔΣ ADCs

2.1 Quantization noise self coupling (QNSC)

The proposed ΔΣ modulator architecture with quantization noise self coupling is shown in Fig. 1. The quantization noise is coupled within a single modulator loop. To emphasize a low-distortion modulator architecture, a direct input feedforward path to the quantizer input is inserted. As illustrated in the figure, the quantization noise is delayed by one clock cycle and subtracted from the summing node at the quantizer input. The injected quantization noise raises the modulator order by one. If *STF* is the signal transfer function and *NTF* is the noise transfer function without noise coupling, then the output of the proposed modulator is

$$V(z) = STF(z) \cdot U(z) + NTF(z) \cdot (1 - z^{-1}) \cdot Q(z). \tag{1}$$

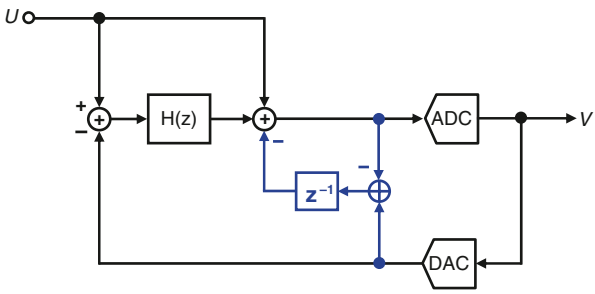


Fig. 1 Noise-coupled modulator with first-order noise coupling

Although the noise coupling within a modulator loop is applicable to any multi-bit modulator, it is even more advantageous to apply it to the low-distortion or feedforward multi-bit modulator architecture, where an extra opamp is usually needed at the quantizer input to realize the active signal summation. Hence, the noise coupling can be implemented without extra dedicated opamp. The stable maximum input amplitude is somewhat reduced since the mean-square value of the injected quantization noise is doubled. The appropriate quantizer resolution should be selected considering this. Generally, it is recommended to select the quantizer resolution (M in bits) to be higher than the effective modulator order (L) by one, i.e. $M=L+1$ to keep the stable maximum input amplitude for multi-bit modulator. On the contrary, non-zero poles are introduced to stabilize the loop in a single-bit modulator [12, 15–18].

Figure 2 shows the equivalent block diagram of the proposed architecture. The extra branch corresponds to a delay-free integrator combined with a delayed DAC at its input. According to the equivalent block diagram, the summing node of the noise-coupled modulator is located at the input of the last integrator rather than of the quantizer. In contrast, the signal summation is performed at the quantizer input in the conventional low-distortion or feedforward ΔΣ ADC, which is one of the disadvantages in realizing this modulator architecture. Thus, noise coupling is not only equivalent to adding more integrators to the loop filter to enhance the NTF, but it also reduces the number of amplifiers needed by advancing the signal summation node to the input of the last integrator in the low-distortion or feedforward modulator. On the other hand, the DEM signal processing time is not relaxed in the noise-coupled architecture since the last integrator is delay-free. It is relaxed when a complete phase is available for the quantizer and DEM circuitry. This can be achieved by inserting extra delay in the input signal path before the loop filter [19, 20].

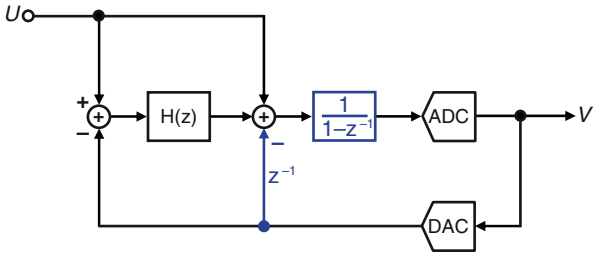


Fig. 2 Equivalent block diagram of Fig. 1

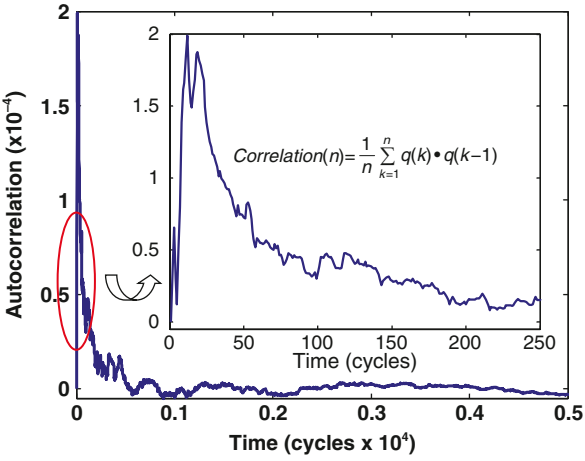


Fig. 3 Time evolution of the autocorrelation of the quantization noise with a one clock period lag for the system of Fig. 1

Using a more sophisticated coupling filter instead of a simple delay in the coupling path, the modulator order can be increased by two or more [21–26]. It can be shown that the transfer function $G(z)$ of a second-order enhancement coupling filter should be $G(z)=2z^{-1}-z^{-2}$. Generally, $G(z)=1-(1-z^{-1})^N$ should be used for an Nth-order enhancement coupling filter. Since the proposed noise-coupled modulator contains the coupling branch following the loop filter, its circuit errors will be attenuated by the preceding loop filter, and hence the related performance improvement is insensitive to these circuit errors.

An additional advantage of the noise-coupled modulator is reduced tone generation and harmonic distortion (THD). This results from the fact that the injected quantization noise works as a random dither signal. Thus, idle tones and harmonic spurs are reduced and the modulator linearity is improved. The role of an injected quantization noise as a random dither is more clearly visualized by observing its time-domain autocorrelation sequence with a one clock cycle delay. Figure 3 shows the simulated time evolution of the autocorrelation sequence $C(n)$ between the quantization noise and its replica delayed by one clock cycle in the noise-coupled modulator with 4-bit quantizer. $C(n)$ is defined as

$$C(n) = \frac{1}{n}\sum_{k=1}^{n} q(k)\cdot q(k-1). \tag{2}$$

This demonstrates that the quantization error is close to white noise in the noise-coupled multi-bit modulator, which is usually not valid in a single-bit modulator. The quantization noise of the proposed modulator gets more randomized for the same input signal condition with multi-bit quantization and noise coupling. The improvement of the modulator linearity is demonstrated in Fig. 4. The figure illustrates the simulated power spectral density of the proposed modulator in the presence of practical circuit errors before and after the noise coupling is applied. The graphs show that the harmonic spurs are reduced by more than 12 dB.

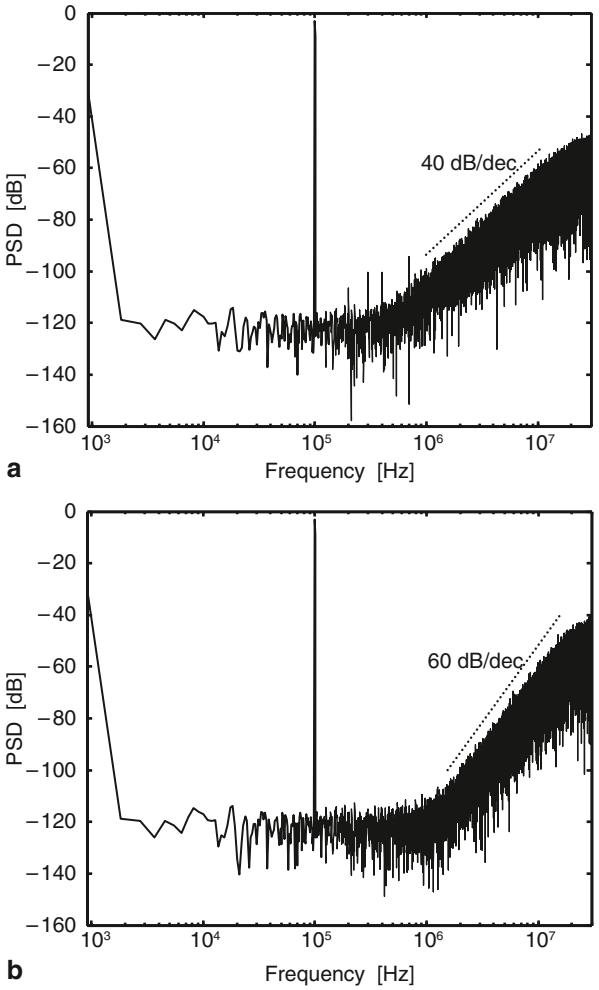


Fig. 4 Simulated power spectral density (PSD) of the modulator including realistic circuit errors; **a** without noise coupling, **b** with noise coupling

Figures 5–7 show a noise-coupled third-order modulator with first-order coupling filter, a conventional third-order modulator, and a noise-coupled third-order modulator with second-order coupling filter, respectively. Each modulator was evaluated by simulation with a realistic non-ideal ΔΣ ADC model in Matlab and Simulink. Figure 8 shows that all three modulators provide nominally identical SNDR performance and input signal saturation level. Each noise coupling filter in Figs. 5 and 7 improves the noise shaping to third order. The noise coupling in Fig. 5 allows over 13 dB SNDR improvement with an OSR of 16. The input saturation level for each third-order modulator is 2 dB lower than that of the second-order modulator (Fig. 5 without noise coupling).

This can be explained by considering the maximum signal amplitude of the summing node in front of the quantizer in a low-distortion feedforward modulator. The summing node has the maximum internal signal swing in this class of modulators.

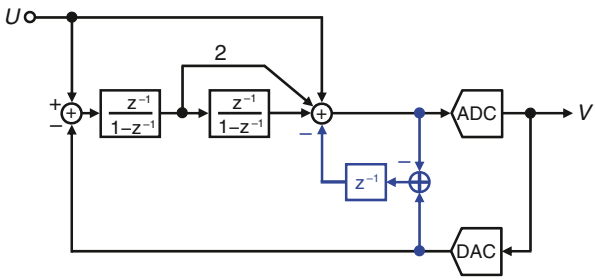


Fig. 5 Third-order noise-coupled modulator with first-order noise coupling

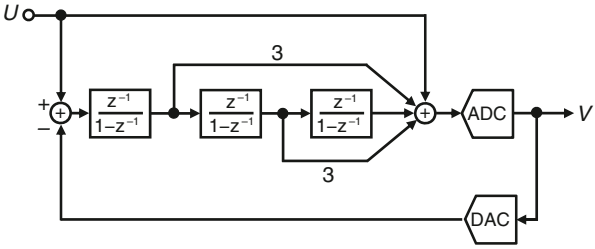


Fig. 6 Conventional third-order low-distortion feedforward modulator

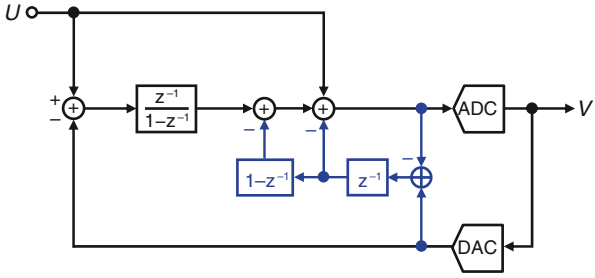


Fig. 7 Third-order noise-coupled modulator with second-order noise coupling

Generally, for a conventional Lth-order low-distortion modulator with an M-bit quantizer, the worst-case maximum signal swing at the summing node is

$$S_{\max} = (2^L - 1) \cdot \Delta_q + \left| V_{in,\max} \right| . \tag{3}$$

Here, Δ_q is the quantizer threshold step

$$\Delta_q = \left| V_{thq}^{+} - V_{thq}^{-} \right| \Big/ 2^M \tag{4}$$

where V_{thq}^{+} and V_{thq}^{-} are the positive and negative reference voltages for the quantizer, respectively. The worst-case signal swing is obtained by assuming complete correlation among shaped or delayed quantization noises and an input signal. Then,

$$S_{\max} = (2^L - 1)/2^M \cdot \left| V_{thq}^{+} - V_{thq}^{-} \right| + \left| V_{in,\max} \right| \tag{5}$$

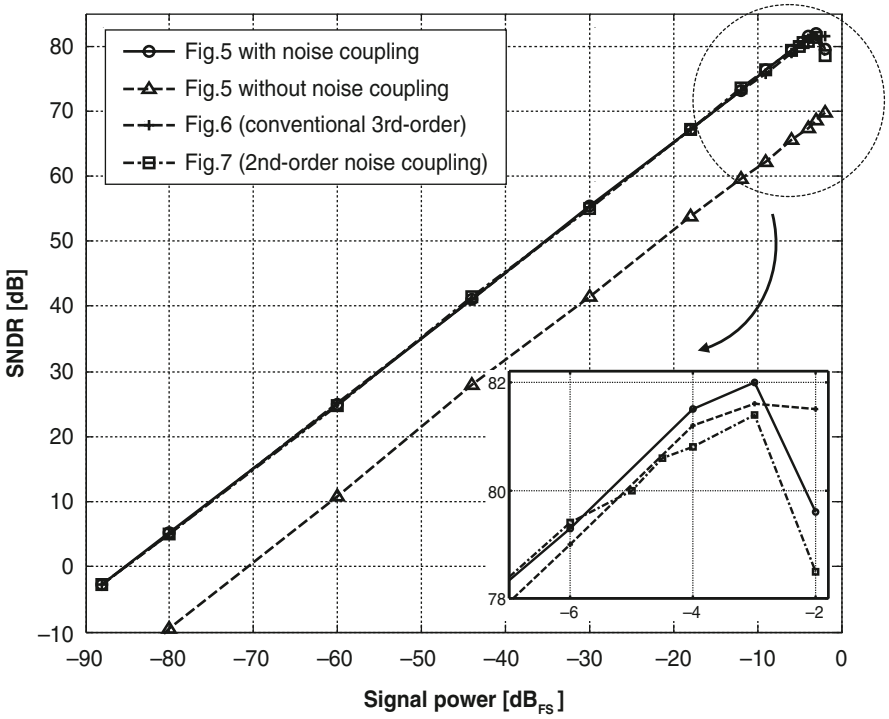


Fig. 8 Simulated SNDR variation with input signal power: top lines for circuits of Figs. 5–7, bottom line for circuit of Fig. 5 without noise coupling

results and it indicates that the stable maximum input signal amplitude can be maintained by selecting its quantizer resolution $M=L+1$ to avoid quantizer overload. Here, the quantizer resolution M of the modulator is defined as $log_2 N_Q$, where N_Q is the number of quantization levels. Since the same quantizer resolution ($M=3.9$ bit) is assumed for all modulators, the signal swing of the summing node in a third-order modulator is larger than that in a second-order one according to Eq. (5). This causes a reduced input signal saturation level for three third-order modulators. Since the injected quantization noise increases the signal swing at the internal nodes of the loop, a proper selection of the quantization levels will provide improved dynamic range and stability [12, 27].

The stability of the noise-coupled modulator is comparable to that of the original modulator with dither, but its dynamic range is improved.

Figure 9 shows the simulated SNDR variation with gain errors in the noise-coupling branch caused by capacitor mismatch, and with offset errors in the quantizer, for the modulator shown in Fig. 5 when a −6 dBFS input signal is applied. Since the noise-coupling capacitor mismatch is typically much less than 5%, and the quantizer offsets can be reduced by input offset sampling scheme, the SNDR degradation

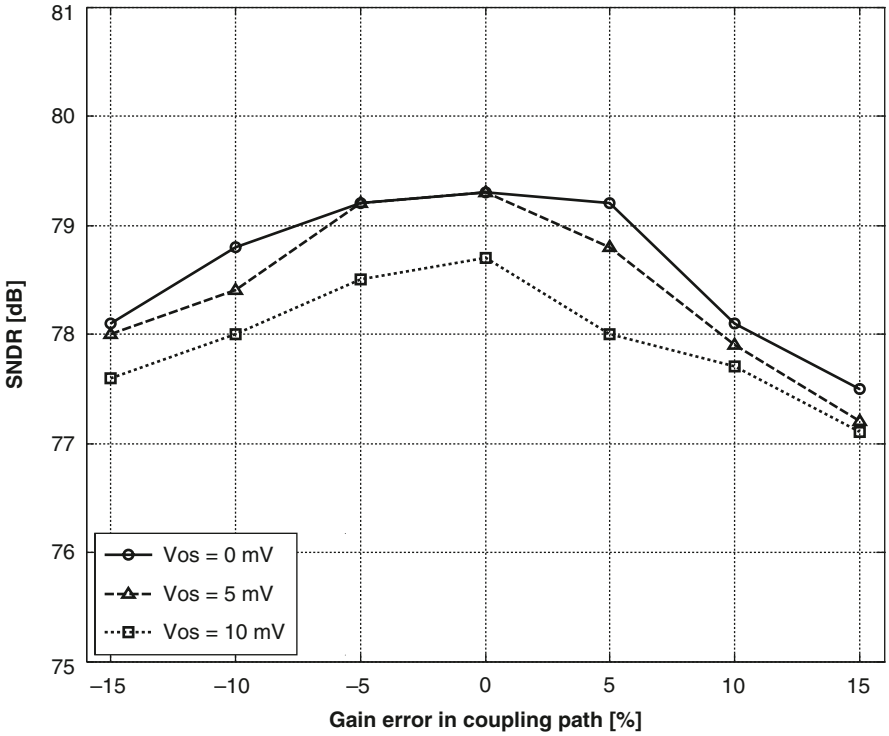


Fig. 9 Simulated SNDR variation with gain errors in the noise-coupling path due to the associated capacitor mismatch and with rms offset errors in the quantizer for the modulator shown in Fig. 4

due to these errors is within 1 dB. The SNDR variation with finite dc gain of active adder opamp is negligible when the gain is higher than 35 dB.

2.2 Quantization noise cross-coupling and time-interleaving (NCTI)

The ΔΣ ADC described in this section uses a noise-coupled time-interleaved architecture. Its derivation is illustrated in Fig. 10. Figure 10a shows a split modulator using low-distortion configuration in each path. The split ADC architecture was proposed earlier in pipeline and successive approximation (SAR) ADCs, to facilitate digital background calibration [28, 29]. In splitting the modulator, all capacitances and transconductances (g_m) are cut in half. For constant power dissipation, the split modulator provides the same SNR and signal bandwidth as the single-path one. This is because both the thermal and quantization noise are in-

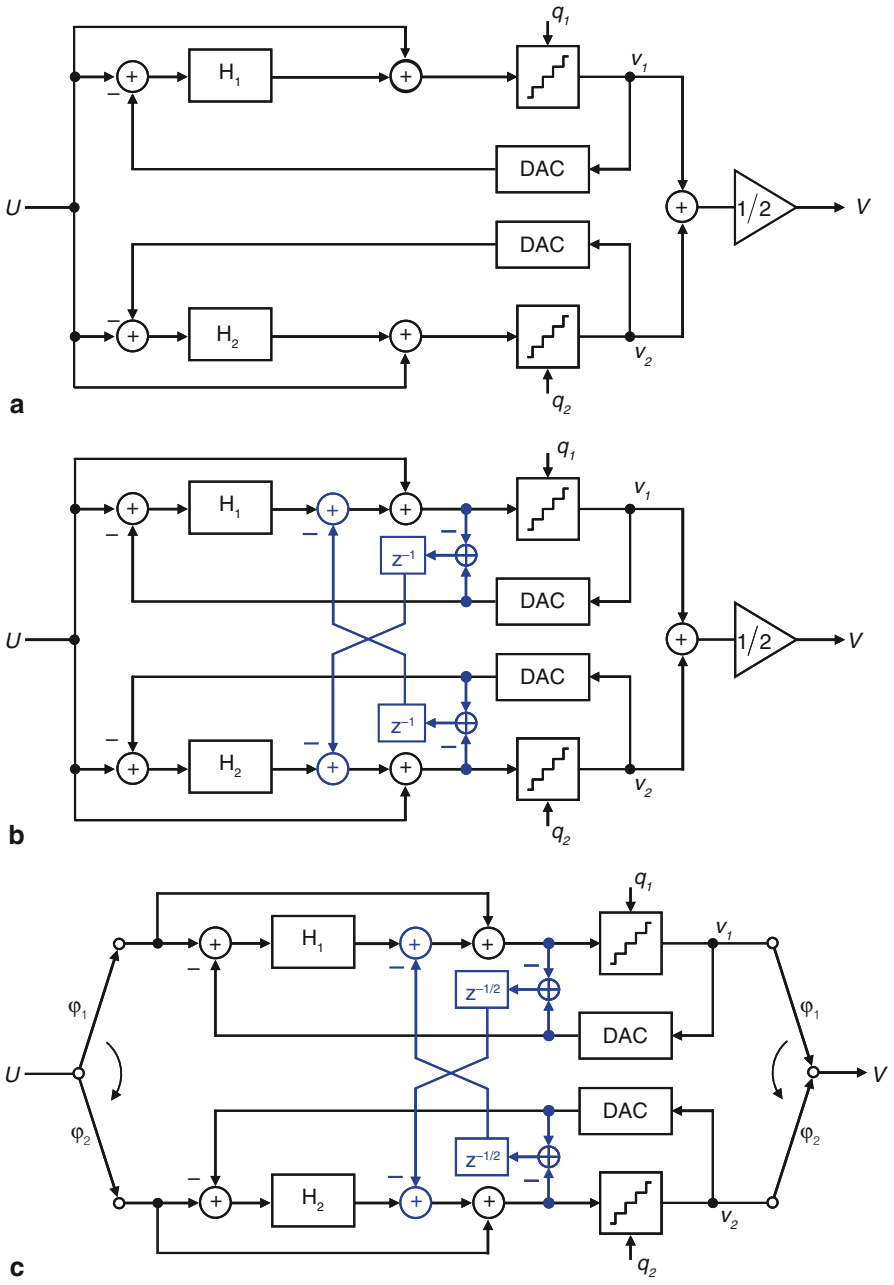


Fig. 10 Derivation of the proposed modulator architecture. **a** split modulator; **b** noise-coupled split (NCS) modulator; **c** noise-coupled time-interleaved (NCTI) modulator

creased by 6 dB, and so is the output signal. The z-domain equations for the split modulator are

$$V_1 = U + NTF_1 \cdot Q_1 \tag{6}$$

$$V_2 = U + NTF_2 \cdot Q_2 \tag{7}$$

$$V = U + NTF \cdot (Q_1 + Q_2)/2. \tag{8}$$

Here, a unity signal transfer function and an identical noise transfer function ($NTF_1 = NTF_2 = NTF$) for each cell modulator are assumed. Equation (8) illustrates that the root-mean-square (rms) power of the quantization noise for the split modulator is reduced by 3 dB compared with that of the single-path one, if the same number of quantization levels is kept for all modulators.

The next step is to introduce quantization noise coupling between the two ΔΣ modulators, as shown in Fig. 10b. The quantization noise of each modulator is delayed by one cycle, and injected into the summing node at the quantizer input of the other modulator. Similar to the previous quantization noise self coupling, this results in a new noise transfer function for the combined modulator [21–26]. The resulting equations are

$$V_1 = U + NTF_1 \cdot (Q_1 - z^{-1} \cdot Q_2) \tag{9}$$

$$V_2 = U + NTF_2 \cdot (Q_2 - z^{-1} \cdot Q_1) \tag{10}$$

$$V = U + NTF \cdot (1 - z^{-1}) \cdot (Q_1 + Q_2)/2. \tag{11}$$

In a split modulator with and without noise coupling, the quantization noises of the two loops are decorrelated quickly (typically within a few hundred cycles after power-up) in the presence of thermal noises and circuit errors such as opamp offsets, settling errors and nonlinearities. A rigorous proof can be found in [30]. A similar plot can be obtained with simulation for the cross-correlation of the quantization noises from the two loops where the cross-correlation $C(n)$ is defined as

$$C(n) = \frac{1}{n} \sum_{k=1}^{n} q_1(k) \cdot q_2(k). \tag{12}$$

Further performance improvement is achieved by adding time interleaving to the system. The resulting noise-coupled time-interleaved (NCTI) modulator is illustrated in Fig. 10c [24, 25]. As shown in the simplified block diagram of Fig. 11a, time interleaving changes the extra factor in the noise transfer function from $(1 - z^{-1})$ to $(1 - z^{-1/2})$. This new factor gives an additional 6 dB SQNR improvement in the signal band. This occurs since $(1 - z^{-1/2}) \sim (1 - z^{-1})/2$ at low frequencies where $z \sim 1$. The related equations under this modification are

$$V_1 = U_1 + NTF_1 \cdot (Q_1 - z^{-1/2} \cdot Q_2) \tag{13}$$

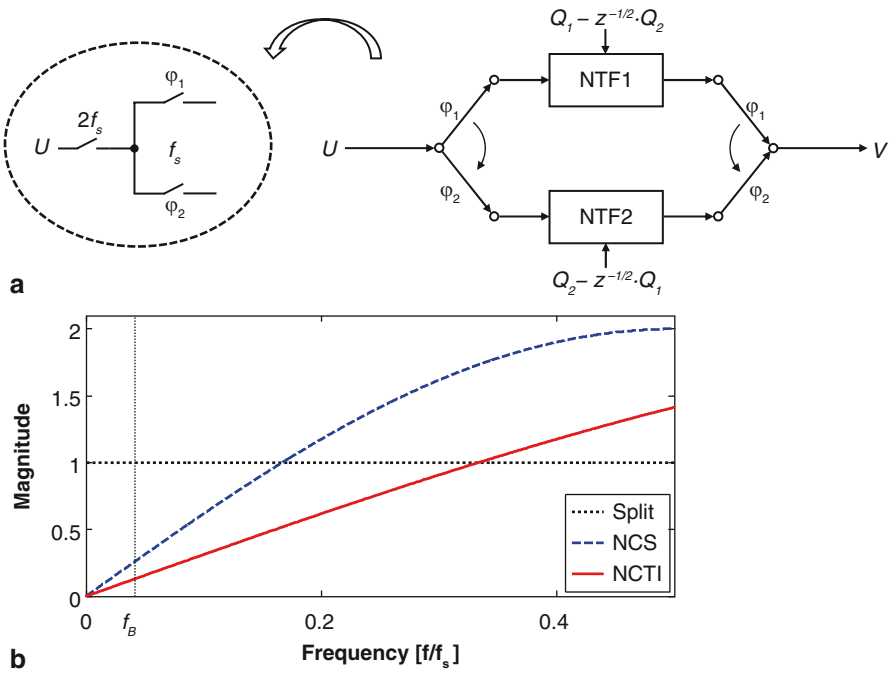


Fig. 11 **a** Block diagram of a NCTI modulator. **b** Comparison of the magnitudes of the NTF enhancement factors for various split modulators (linear scale in both axes); 1, $|1 - z^{-1}|$, and $|1 - z^{-1/2}|$

$$V_2 = U_2 + NTF_2 \cdot (Q_2 - z^{-1/2} \cdot Q_1). \tag{14}$$

Considering the phase difference between the two modulators caused by the time-interleaving operation, all the variables in either of the two modulators (in this example, modulator 2) should be multiplied by $z^{-1/2}$ factor before combining the above two equations as shown below

$$V_2 \rightarrow z^{-1/2} \cdot V_2 \tag{15}$$

$$U_2 \rightarrow z^{-1/2} \cdot U_2 \tag{16}$$

$$Q_2 \rightarrow z^{-1/2} \cdot Q_2 \tag{17}$$

Hence, the above two equations are modified to

$$V_1 = U_1 + NTF_1 \cdot (Q_1 - z^{-1/2} \cdot z^{-1/2} \cdot Q_2) \tag{18}$$

$$z^{-1/2} \cdot V_2 = z^{-1/2} \cdot U_2 + NTF_2 \cdot (z^{-1/2} \cdot Q_2 - z^{-1/2} \cdot Q_1). \tag{19}$$

Combining these equations using the following definitions

$$V = V_1 + z^{-1/2} \cdot V_2 \tag{20}$$

$$U = U_1 + z^{-1/2} \cdot U_2 \tag{21}$$

$$Q = Q_1 + z^{-1/2} \cdot Q_2 \tag{22}$$

we get the final expression

$$V = U + NTF \cdot (1 - z^{-1/2}) \cdot Q. \tag{23}$$

As in a NCS modulator, the quantization noises of the two loops in a NCTI modulator become uncorrelated within a few hundred cycles after power-up. The signal of a NCTI modulator increases by about a factor of 2. This is due to the fact that U_1 and U_2 are strongly correlated and $|1+z^{-1/2}|$ is very close to 2 in the signal band of an oversampling ADC, unlike in a Nyquist sampling ADC. For example, $|u_1(n) - u_2(n)|$ is less than 6.6% of the amplitude of $u_1(n)$ and $u_2(n)$ in the signal band, and $|1+z^{-1/2}|=1.9957$ when OSR = 12. Thus,

$$U = U_1 + z^{-1/2} \cdot U_2 \approx (1 + z^{-1/2}) \cdot U_1 \approx 2 \cdot U_1. \tag{24}$$

For a fair comparison, if we scale the signals of a NCTI modulator to have the same signal power as in a split modulator and an NCS one, we obtain

$$V/2 \approx U_1 + NTF \cdot (1 - z^{-1/2}) \cdot (Q_1 + z^{-1/2} \cdot Q_2)/2. \tag{25}$$

Comparing Eqs. (8), (11), and (25), we see that the performance improvement of a NCTI modulator comes from the enhancement factor $(1 - z^{-1/2})$, since both $(Q_1+Q_2)/2$ and $(Q_1+z^{-1/2} \cdot Q_2)/2$ give the same noise power once Q_1 and Q_2 are decorrelated. The frequency responses of the enhancement factors of the three modulators shown in Fig. 10 are compared in Fig. 11b. Time interleaving also allows a simpler realization of the noise coupling branches since the $z^{-1/2}$ delay is automatically realized when injecting the quantization of one modulator into the other. Finally, the NCTI modulator has reduced sensitivity to channel mismatch errors, as a result of the suppressed quantization noise around the Nyquist frequency due to the out-of-band zeros of the modulator.

Generally, a time-interleaved Nyquist ADC suffers from channel mismatch errors, such as timing skew, channel gain mismatch, channel offset mismatch, and channel bandwidth mismatch [31–33]. Foreground or background calibration is often required, either in the analog or digital domain, to reduce these errors [34, 35]. By contrast, channel mismatch errors do not limit the performance of the proposed NCTI modulator. The robust performance to channel mismatches is due to the out-of-band NTF zeros of the NCTI modulator at the Nyquist frequency f_s (Fig. 12). A double zero is introduced by operating the two second-order ΔΣ modulators with

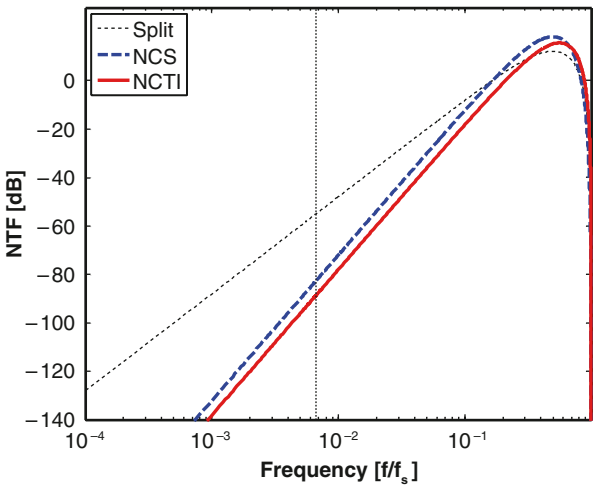


Fig. 12 Comparison of the NTFs of the three modulators shown in Fig. 10, with a second-order loop filter $H(z)$ (logarithmic scale in both axes)

time-interleaved sampling. Since there is thus no significant quantization noise power at the Nyquist frequency f_s, the modulator can tolerate noise folding due to channel mismatches up to 5%. However, the added double zero at the Nyquist frequency also raises the in-band noise floor, exhibiting a trade-off between SQNR performance and robustness to channel mismatch errors.

The proposed modulator topology was realized with second-order low-distortion loops. Figure 12 shows the NTF magnitude responses of the three ΔΣ ADCs introduced in Fig. 10. The dotted line for the uncoupled second-order modulator shows a 40 dB/decade slope. The dashed line for the NCS third-order modulator shows a 60 dB/decade slope. Finally, the NCTI modulator has the same slope as the NCS one, but the curve is 6 dB lower. With an OSR of 12, the SQNR improvement with the NCTI modulator is 19 dB, as compared to a conventional second-order modulator, and it requires only some additional switched capacitors. A realistic ΔΣ modulator model was simulated with Matlab and Simulink to verify the concept. Figure 13 shows the

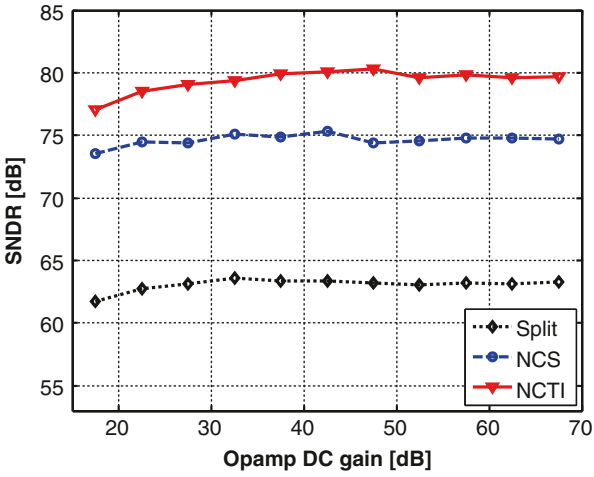


Fig. 13 SNDR variation with opamp DC gain whose absolute mismatch is 5% between the two channels

SNDR as a function of the integrator opamp dc gains and channel mismatch. The proposed modulator shows more than 17 dB SNDR improvement over the conventional modulator, and its performance is robust in the presence of various circuit and channel mismatch errors.

3 SC circuit implementation

Several prototype ΔΣ ADCs were implemented in a 0.18 μm 2P4M CMOS process. The block diagram of the realized ΔΣ ADC with self noise-coupling is illustrated in Fig. 14; its simplified switched-capacitor (SC) circuit diagram in Fig. 15. The block diagram and SC realization for NCTI ΔΣ ADC are shown in Figs. 16 and 17. The details of the input circuitry are omitted here for simplicity. The detailed input branch for the first integrator for both prototypes is shown in Fig. 18. The input sampling and DAC capacitors are separated to avoid detrimental reference errors caused by signal-dependent loading of the reference driver [36]. However, the separation of the input sampling and DAC capacitors slows down the settling of the opamp, due to a reduction of the feedback factor of the opamp. The input sampling and DAC capacitors were reduced by half to preserve the feedback factor, and the signal swings for input and DAC reference are also doubled to keep the same SNR using cross-coupled switches between positive and negative rails, as illustrated in Fig. 18.

Each loop contains two integrators, followed by an active adder. The low-distortion modulator requires signal summation at the input of the quantizer. Passive SC adder, which does not require extra opamp, could be used, but the signal swing is then reduced due to parasitic capacitances, especially when the loop contains a multibit quantizer. The SC adder is also susceptible to kick-back noise from the quantizer. For these reasons, active rather than passive summing with an offset-compensated amplifier was used in this design. The noise coupling also utilized

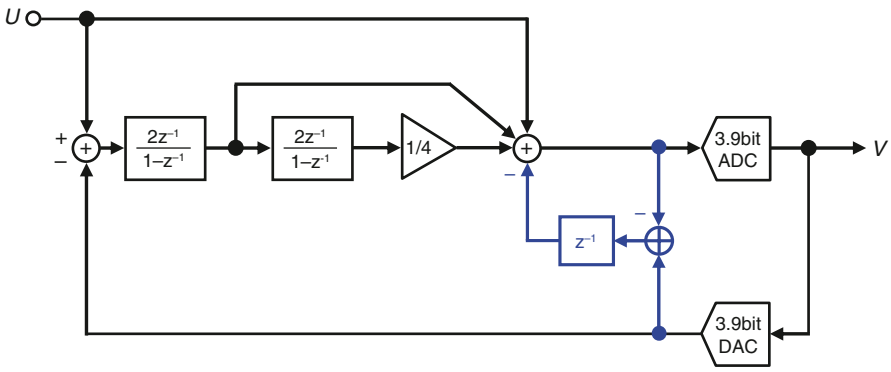
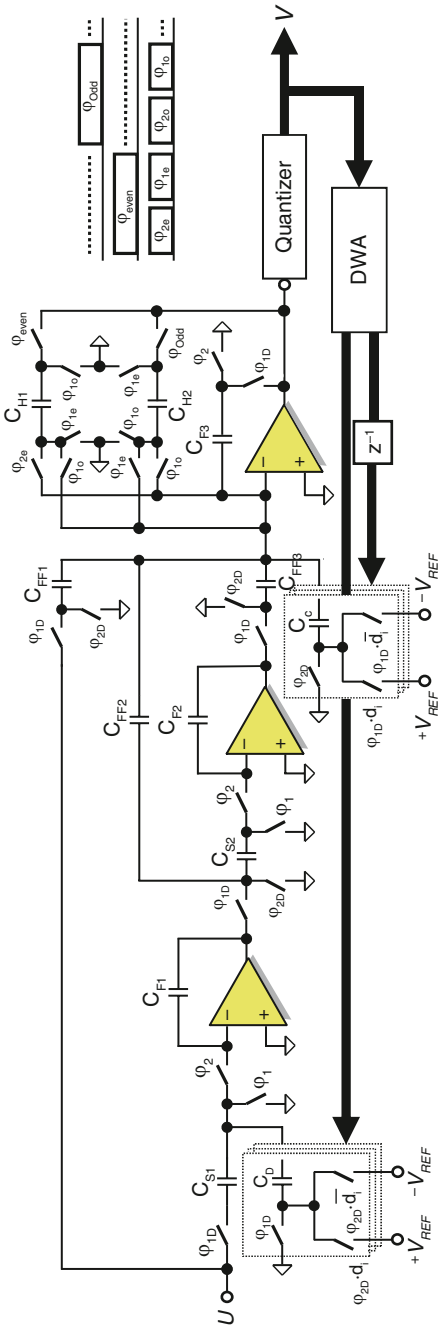


Fig. 14 Block diagram of the QNSC prototype ΔΣ ADC



	φ_{2e}	φ_{1e}	φ_{2o}	φ_{1o}
C_{H1}	Reset	Sample	Float	Transfer
C_{H2} z^{-1}	Float	Transfer	Reset	Sample
C_{F3}	Reset	Transfer	Reset	Transfer

Fig. 15 SC circuit implementation of the prototype ΔΣ ADC in Fig. 14

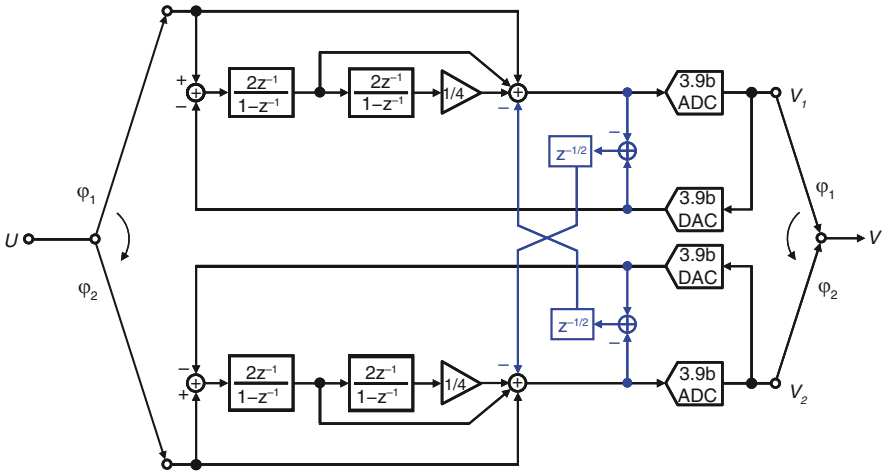


Fig. 16 Block diagram of the NCTI prototype ΔΣ ADC

the active adder opamp. To compensate for the low OSR of 12, 15-level quantizers were used along with 15-level unit-capacitor DACs.

The signal swing of the loop filter is significantly reduced, thanks to the low-distortion modulator architecture and multi-bit quantization. As a result, the limited output signal swing of the opamp was not an issue even with a 1.5 V supply. This allows the use of the power-efficient and low-noise telescopic cascode opamp with a SC common-mode feedback, as illustrated in Fig. 19, for the integrators and the active adder. The differential output signal swing of the opamp is ±0.72 V. The slew rate requirement for the opamp is also greatly relaxed by using the low-distortion topology and multi-bit quantization [12].

The required low GBW and slew rate requirements of the opamp were confirmed by system-level simulation with Matlab and Simulink before its circuit realization. The first integrator opamp in NCTI modulator achieves a GBW of 1.2 GHz and a DC gain of 50 dB, while drawing 2.6 mA bias current. Similar but scaled opamps were used for the second integrator and the active adder. An offset compensation scheme was employed in the delay-free active adder. The critical input sampling switches use bootstrapped clock signals to insure linear sampling [37].

The 4b quantizer contains a two-stage preamplifier, to prevent kick-back noise and to speed up signal amplification, followed by a track-and-latch and a set-and-reset (SR) latch as shown in Fig. 20 [38, 39]. Input offset sampling scheme is employed in the two-stage preamplifier to minimize the effects of offset errors. Since the preamplifier has an open-loop gain of around 10, the input-referred offset is reduced by about this factor. During the comparison phase φ_1, the track-and-latch is in the track mode and senses the output of the preamplifier. Before the end of φ_1 phase, the φ_c (reset) signal of the track-and-latch stage goes from high to low, and triggers the regeneration and latching. The threshold voltages of the quantizer are generated by a resistor string.

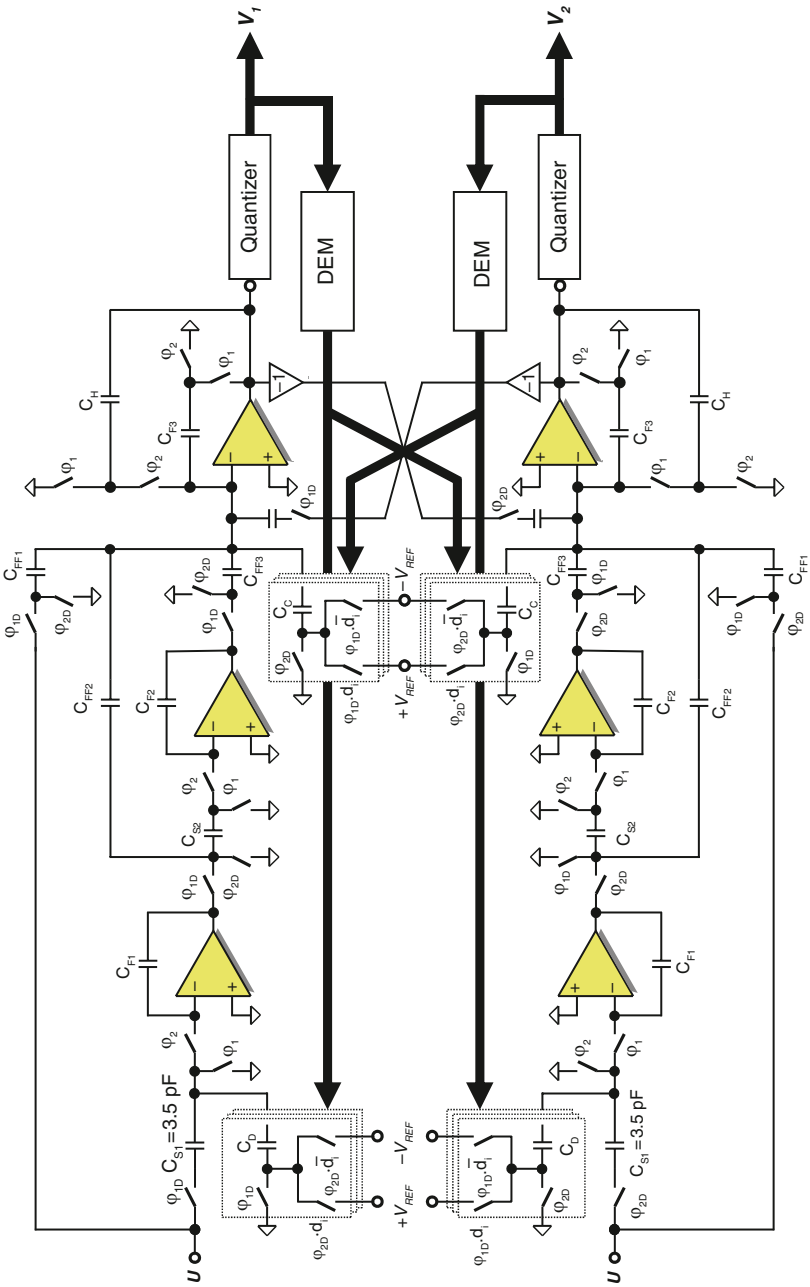


Fig. 17 SC circuit implementation of the prototype ΔΣ ADC in Fig. 16

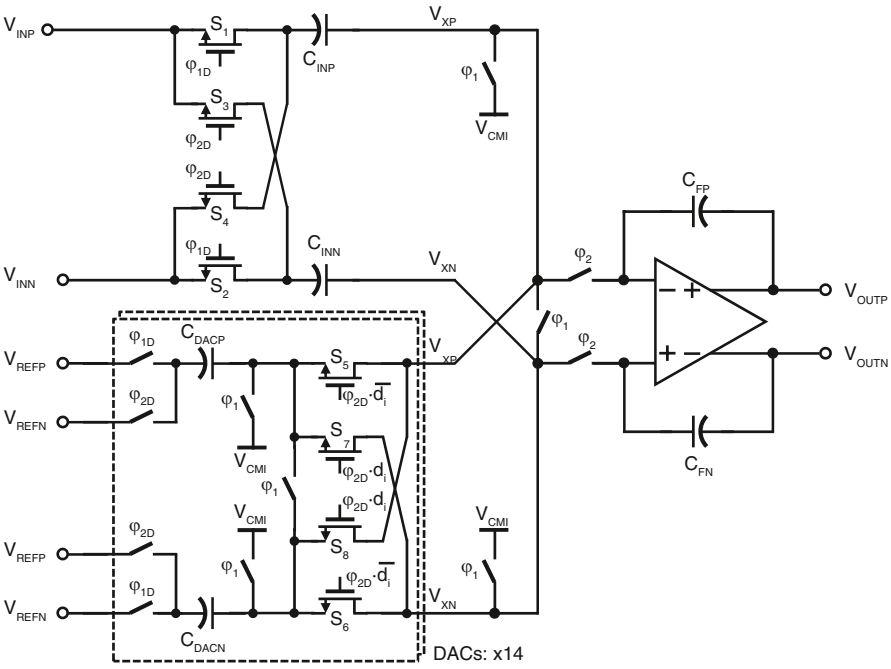


Fig. 18 Detailed input branch of the modulator with separate input and DAC capacitors

Conventional data-weighted averaging (DWA) was applied to shape the mismatch errors of the 15-level DAC [40]. Figure 21 shows the block diagram of the DWA implementation. The transistors in the quantizer and in the four-stage logarithmic shifter on the critical path were optimized in size to meet the critical timing requirements.

4 Experimental Results

The prototype devices were fabricated in a 0.18 μm 2P4M CMOS process.

4.1 QNSC prototype ΔΣ ADC

The measured output spectrum for QNSC prototype ADC is illustrated in Fig. 22. The SNR and SNDR variations with input signal power are shown in Fig. 23. The prototype ADC achieves 81 dB peak SNDR, 82 dB DR, and −98 dB THD in the 1.9 MHz signal band. The measured total power dissipation is 8.1 mW (analog:

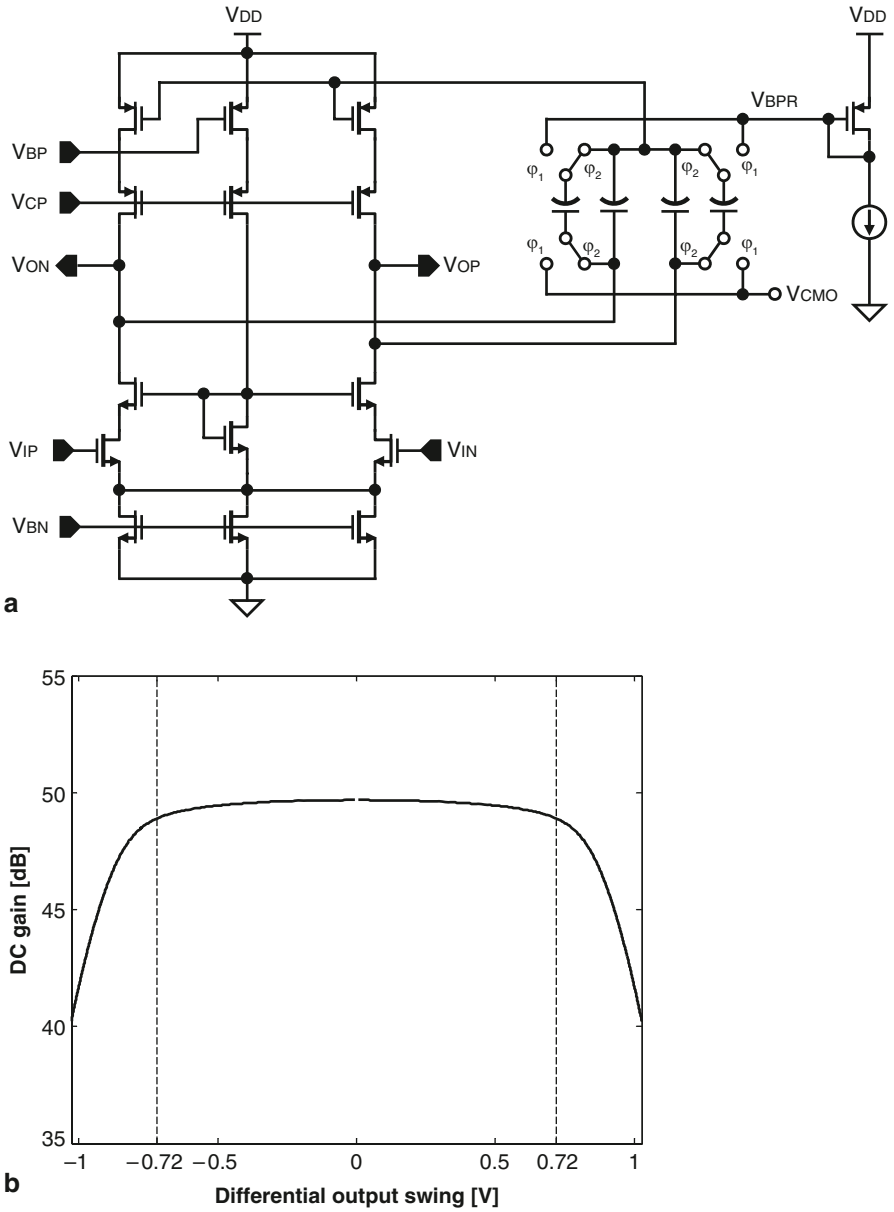


Fig. 19 **a** Telescopic cascode opamp with SC common-mode feedback. **b** simulated dc gain vs. differential output swing

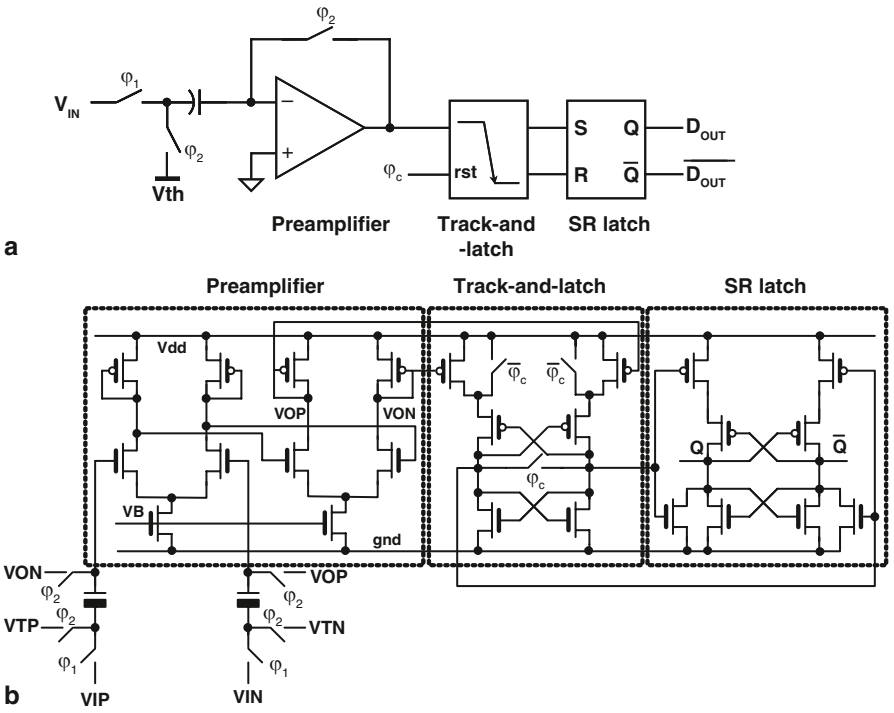


Fig. 20 Schematic of quantizer: **a** block diagram, **b** circuit implementation

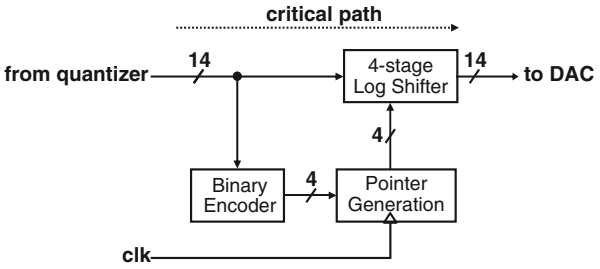


Fig. 21 A simplified block diagram of DWA using a 4-stage logarithmic shifter

4.4 mW, digital: 3.7 mW). The measured performance is summarized in Table 1. The figure-of-merit (FOM), defined by $\mathrm{FOM}=\mathrm{P}/(2\cdot\mathrm{BW}\cdot 2^{(\mathrm{SNDR}-1.76)/6.02})$, is 0.25 pJ/conversion-step, which is among the lowest published for wideband DT $\Delta\Sigma$ ADCs.

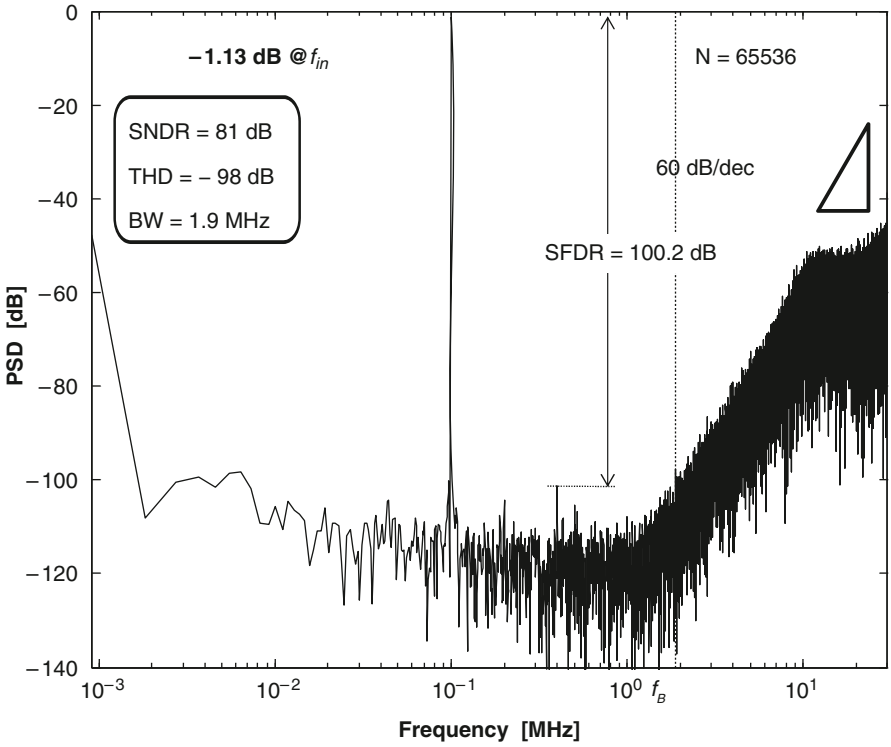


Fig. 22 Measured output spectrum of QNSC prototype ΔΣ ADC

Table 1 Measured performance summary

Architecture	QNSC	NCTI	
Prototype		A	B
Sampling frequency	60 MHz	200 MHz	120 MHz
ΔΣ clock frequency	60 MHz	100 MHz	60 MHz
Signal bandwidth	1.9 MHz	4.2 MHz	2.5 MHz
OSR	16	12	
Input range (diff.)	1.44 V_{pp}	1.44 V_{pp}	
V_{ref}	0.72 V	0.72 V	
C_{IN} and C_{DAC}	separate	separate	shared
Dynamic range	82 dB	81 dB	83 dB
SNDR	81 dB	79 dB	81 dB
THD	−98 dB	−98 dB	−104 dB
FOM	0.25 pJ/conv.	0.48 pJ/conv.	0.33 pJ/conv.
Power consumption	4.4 mW (A) 3.7 mW (D)	13 mW (A) 15 mW (D)	10 mW (A) 5 mW (D)
Power supply	1.5 V (A), 1.45 V (D)	1.5 V (A), 1.6 V (D)	
Process	0.18 µm 2P4M CMOS	0.18 µm 2P4M CMOS	
Core area	1.27 mm^2	3.67 mm^2	

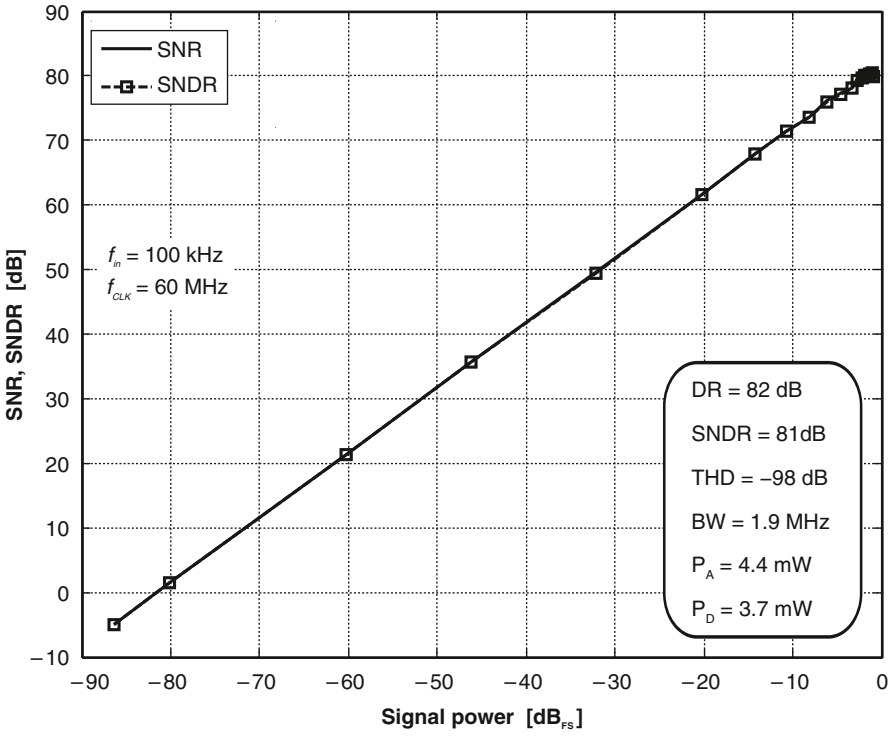


Fig. 23 Measured SNR and SNDR with input signal power for QNSC prototype ΔΣ ADC

4.2 NCTI prototype ΔΣ ADCs

Two types of NCTI prototype ADCs were fabricated and measured. Both prototype ADCs are identical except for their input sampling branches. Prototype A has separate input sampling and DAC capacitors and prototype B has shared input sampling and DAC capacitors. For prototype A, Fig. 24 shows the measured output spectrum and Fig. 25 shows the SNR and SNDR variations with input signal amplitude. Prototype A achieves 79 dB peak SNDR, 81 dB DR, and −98 dB THD in the 4.2 MHz signal band. The analog power is 13 mW and the digital power is 15 mW. The FOM is 0.48 pJ/conversion-step. Prototype B shows 81 dB peak SNDR, 83 dB DR, and −104 dB THD in the 2.5 MHz signal band. The analog power is 10 mW and digital power is 5 mW. The FOM is 0.33 pJ/conversion-step. Table 1 summarizes the measured performance for both prototypes. The FOMs of all three prototype ADCs are compared with other state-of-the-art ΔΣ ADCs in Fig. 26.

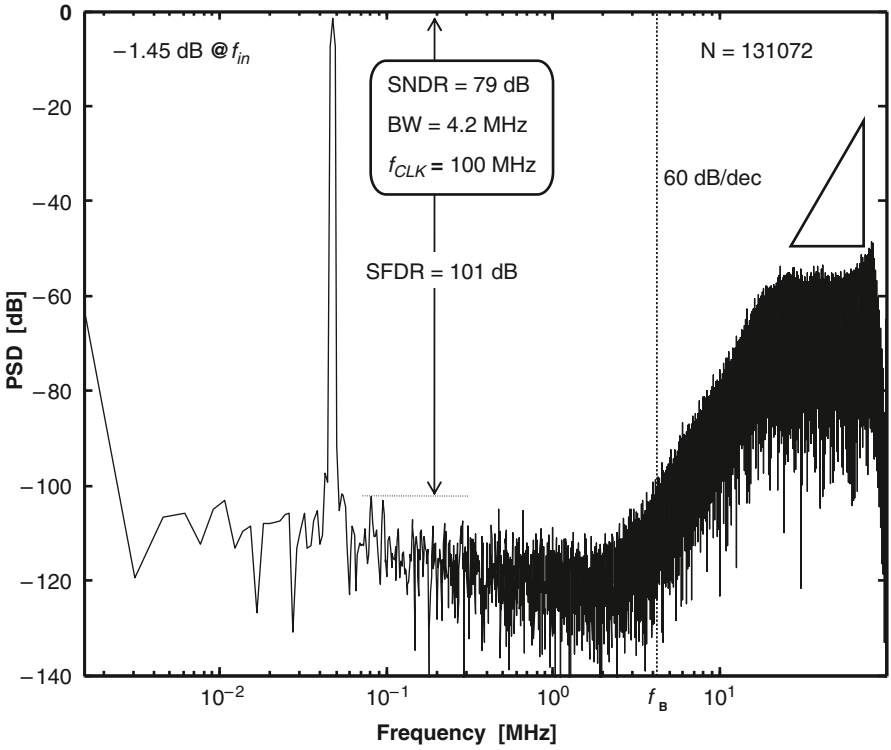


Fig. 24 Measured output spectrum of NCTI ΔΣ ADC (prototype A)

5 Conclusions

Highly linear and power-efficient wideband DT ΔΣ ADCs were described. Their superior performance was achieved with effective architectural improvement and circuit implementation. The proposed modulator architecture are based on the novel noise coupling and time interleaving. The first prototype ADC is an improved low-distortion ΔΣ modulator using a internal noise coupling. It requires fewer opamps than a conventional low-distortion modulator of the same order, by sharing the active adder opamp in front of the quantizer with a noise coupling branch. The prototype ADC shows highly linear data conversion (−98 dB THD) and state-of-the-art power efficiency (FOM = 0.25 pJ/conversion-step) over a 1.9 MHz signal band. The ADC's power efficiency is among the best for recently reported DT and CT ΔΣ modulators. The second prototype ADCs are noise-coupled time-interleaved ΔΣ ADCs. The advantages with these modulators are enhanced noise shaping and ro-

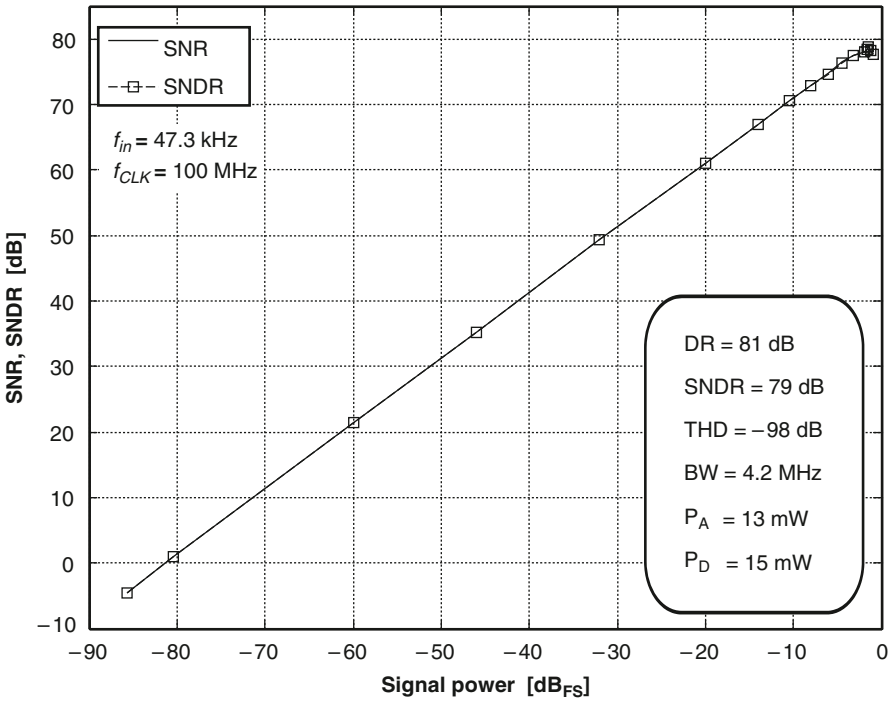


Fig. 25 Measured SNR and SNDR with input signal power for NCTI ΔΣ ADC (prototype A)

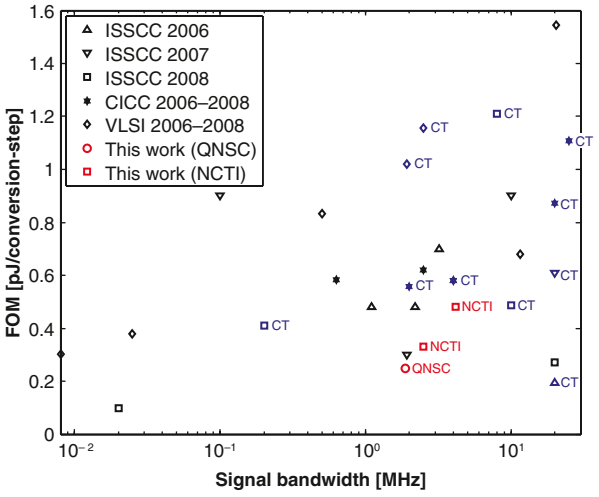


Fig. 26 Comparison with other state-of-the-art ΔΣ modulators (CT and the QNSC and NCTI modulators of this work are marked. All other modulators are DT)

bust performance. This makes them suitable for wideband low-power modulators. In addition, the coupled noise reduces the THD to a level lower than that normally achieved in broadband ADCs. The measured performance of a prototype device showed a 79 dB SNDR and −98 dB THD in a 4.2 MHz signal band. This compares well with the theoretical maximum SQNR for a conventional third-order modulator with a 4b quantizer, which is 80 dB with an OSR of 12.

References

1. T. Christen, T. Burger, Q. Huang, A 0.13 µm CMOS EDGE/UMTS/WLAN Tri-Mode ΔΣ ADC with −92 dB THD, in *Proc. IEEE ISSCC Dig. Tech. Papers*, San Francisco, Feb 2007, pp. 240, 241
2. S. Kwon, F. Maloberti, A 14 mW multi-bit ΔΣ modulator with 82 dB SNR and 86 dB DR for ADSL2+, in *Proc. IEEE ISSCC Dig. Tech. Papers*, San Francisco, Feb. 2006, pp. 161, 162
3. K.-S. Lee, S. Kwon, F. Maloberti, A power-efficient two-channel time-interleaved ΣΔ modulator for broadband applications, *IEEE J. Solid-State Circuits.* **42**(6), 1206–1215 (2007)
4. Y. Kanazawa, Y. Fujimoto, P. Lo Re, M. Miyamoto, A 100-MS/s 4-MHz bandwidth 77.3-dB SNDR ΔΣ ADC with a triple sampling technique, in *Proc. IEEE Custom Integrated Circuits Conf.* (*CICC*), San Jose, Sep. 2006, pp. 53–56
5. L. Dorrer, F. Kuttner, P. Greco, P. Torta, T. Hartig, A 3-mW 74-dB SNR 2-MHz continuous-time delta-sigma ADC with a tracking ADC quantizer in 0.13-µm CMOS, *IEEE J. Solid-State Circuits*, **40**(12), 2416–2427 (2005)
6. M. Park, M.H. Perrott, A 78 dB SNDR 87 mW 20 MHz bandwidth continuous-time ΔΣ ADC with VCO-based integrator and quantizer implemented in 0.13 µm CMOS, *IEEE J. Solid-State Circuits*, **44**(12), 3344–3358 (2009)
7. O. Rajaee, T. Musah, S. Takeuchi, M. Aniya, K. Hamashita, P. Hanumolu, U. Moon, A 79 dB 80 MHz 8X-OSR hybrid delta-sigma/pipeline ADC, *IEEE Symp. on VLSI Circuits*, Kyoto, Jun. 2009, pp. 74, 75
8. J.A. Cherry, W.M. Snelgrove, Clock jitter and quantizer metastability in continuous-time delta-sigma modulators, *IEEE Trans. Circuits Syst. II*, **46**(6), 661–676 (1999)
9. G. Mitteregger, C. Ebner, S. Mechnig, T. Blon, C. Holuigue, E. Romani, A 20-mW 640-MHz CMOS continuous-time ΣΔ ADC with 20-MHz signal bandwidth, 80-dB dynamic range and 12-bit ENOB, *IEEE J. Solid-State Circuits*, **41**(12), 2641–2649 (2006)
10. S. Paton, A.D. Giandomenico, L. Hernandez, A. Wiesbauer, T. Potscher, M. Clara, A 70-mW 300-MHz CMOS continuous-time ΣΔ ADC with 15-MHz bandwidth and 11 bits of resolution, *IEEE J. Solid-State Circuits*, **39**(7), 1056–1063 (2004)
11. X. Chen, Y. Wang, Y. Fujimoto, P. Lore, Y. Kanazawa, J. Steensgaard, G.C. Temes, A 18 mW CT ΔΣ modulator with 25 MHz bandwidth for next generation wireless applications, in *Proc. IEEE Custom Integrated Circuits Conf.* (*CICC*), San Jose, Sep. 2007, pp. 73–76
12. R. Schreier, G.C. Temes, *Understanding Delta-Sigma Data Converters* (Wiley, New York, 2004)
13. J. Silva, U. Moon, J. Steensgaard, G.C. Temes, Wideband low-distortion ΔΣ ADC topology, *Electron. Lett.*, **37**(12), 737, 738 (2001)
14. A.A. Hamoui, K.W. Martin, High-order multibit modulators and pseudo data-weighted-averaging in low-oversampling ΔΣ ADCs for broadband applications, *IEEE Trans. Circuits Syst. I, Reg. Papers*, **51**(1), 72–85 (2004)
15. K.C.-H. Chao, S. Nadeem, W.L. Lee, C.G. Sodini, A higher order topology for interpolative modulators for oversampling A/D converters, *IEEE Trans. Circuits Syst.*, **37**(3), 309–318 (1990)

16. R.W. Adams, P.F. Ferguson, A. Ganesan, S. Vincelette, A. Volpe, R. Libert, Theory and practical implementation of a fifth-order sigma-delta A/D converter, *J. Audio Eng. Soc.*, **39**(7/8), 515–528 (1991)
17. R. Schreier, An empirical study of high-order single-bit delta-sigma modulators, *IEEE Trans. Circuits Syst. II, Analog Digit. Signal. Process.*, **40**(8), 461–466 (1993)
18. L. Risbo, Stability predictions for high-order Σ-Δ modulators based on quasilinear modeling, in *Proc. IEEE ISCAS 1994*, London, May 1994, vol. 5, pp. 361–364
19. Y. Fujimoto, Y. Kanazawa, P. Lo Re, M. Miyamoto, An 80/100 MS/s 76.3/70.1 dB SNDR ΔΣ ADC for digital TV receivers, in *Proc. IEEE ISSCC Dig. Tech. Papers*, San Francisco, Feb. 2006, pp. 76, 77
20. K. Lee, G.C. Temes, Improved low-distortion ΔΣ ADC topology, in *Proc. IEEE ISCAS 2009*, Taipei, May 2009, pp. 1341–1344
21. K. Lee, M. Bonu, G.C. Temes, Enhanced split-architecture delta-sigma ADC, *Electron. Lett.*, **42**(13), 737, 738 (2006)
22. K. Lee, M. Bonu, G.C. Temes, Noise-coupled delta-sigma ADCs, *Electron. Lett.*, **42**(24), 1381, 1382 (2006)
23. K. Lee, G.C. Temes, Enhanced split-architecture delta-sigma ADC, in *Proc. IEEE ICECS 2006*, Nice, Dec. 2006, pp. 427–430
24. K. Lee, G.C. Temes, F. Maloberti, Noise-coupled multi-cell delta-sigma ADCs, in *Proc. IEEE ISCAS 2007*, New Orleans, May 2007, pp. 249–252
25. K. Lee, J. Chae, M. Aniya, K. Hamashita, K. Takasuka, S. Takeuchi, G.C. Temes, A noise-coupled time-interleaved delta-sigma ADC with 4.2 MHz bandwidth, −98 dB THD, and 79 dB SNDR, *IEEE J. Solid-State Circuits*, **43**(12), 2601–2612 (2008)
26. K. Lee, M.R. Miller, G.C. Temes, An 8.1 mW, 82 dB delta-sigma ADC with 1.9 MHz BW and −98 dB THD, *IEEE J. Solid-State Circuits*, **44**(8), 2202–2111 (2009)
27. R. Schreier, Y. Yang, Stability tests for single-bit sigma-delta modulators with second-order FIR noise transfer functions, in *Proc. IEEE ISCAS 1992*, San Diego, May 1992, vol. 3, pp. 1316–1319
28. J. Li, G.-C. Ahn, D.-Y. Chang, U.-K. Moon, A 0.9 V 12 mW 5 MSPS algorithmic ADC with 77 dB SFDR, *IEEE J. Solid-State Circuits*, **40**(4), 960–969 (2005)
29. J. McNeill, M. Coln, B. Larivee, "Split-ADC" Architecture for Deterministic Digital Background Calibration of a 16b 1MS/s ADC, *IEEE ISSCC Dig. Tech. Papers*, San Francisco, Feb. 2005, pp. 276–278
30. S. Pamarti, A theoretical study of the quantization noise in split delta-sigma ADCs, *IEEE Trans. Circuits Syst. I, Reg. Papers*, **55**(5), 1267–1278 (2008)
31. N. Kurosawa, H. Kobayashi, K. Maruyama, H. Sugawara, K. Kobayashi, Explicit analysis of channel mismatch effects in time-interleaved ADC systems, *IEEE Trans. Circuits Syst. I, Reg. Papers*, **48**(3), 261–271 (2001)
32. C.S.G. Conroy, D.W. Cline, P.R. Gray, An 8b 85MS/s parallel pipeline A/D converter in 1 μm CMOS, *IEEE J. Solid-State Circuits*, **28**(4), 447–455 (1993)
33. W.C. Black Jr., D.A. Hodges, Time interleaved converter arrays, *IEEE J. Solid-State Circuits*, **15**(12), 1022–1029 (1980)
34. K.C. Dyer, D. Fu, S.H. Lewis, P.J. Hurst, An analog background calibration technique for time-interleaved analog-to-digital converters, *IEEE J. Solid-State Circuits*, **33**(12), 1912–1919 (1998)
35. D. Fu, K.C. Dyer, S.H. Lewis, P.J. Hurst, A digital background calibration technique for time-interleaved analog-to-digital converters, *IEEE J. Solid-State Circuits*, **33**(12), 1904–1911 (1998)
36. E. Fogleman, J. Welz, I. Galton, An audio ADC delta-sigma modulator with 100 dB SINAD and 102 dB DR using a second-order mismatch-shaping DAC, in *Proc. IEEE Custom Integrated Circuits Conf.* (*CICC*), Orlando, Sep. 2000, pp. 17–20

37. M. Dessouky, A. Kaiser, Very low-voltage digital-audio ΔΣ modulator with 88-dB dynamic range using local switch bootstrapping, *IEEE J. Solid-State Circuits*, **36**(3), 349–355 (2001)
38. I. Mehr, D. Dalton, A 500-MSample/s, 6-bit Nyquist-rate ADC for disk-drive read-channel applications, *IEEE J. Solid-State Circuits*, **34**(7), 912–920 (1999)
39. I. Mehr, L. Singer, A 55-mW, 10-bit, 40-Msample/s Nyquist-rate CMOS ADC, *IEEE J. Solid-State Circuits*, **35**(3), 318–325 (2000)
40. R.T. Baird, T.S. Fiez, Linearity enhancement of multibit ΔΣ A/D and D/A converters using data weighted averaging, *IEEE Trans. Circuits Syst. II, Analog Digit. Signal. Process.*, **42**(12), 753–762 (1995)

Very Low OSR Sigma-Delta Converters

Trevor C. Caldwell

1 Introduction

High-speed data converters operating on input signal bandwidths in the megahertz range are key analog building blocks for a variety of applications including high-speed wireless and wireline communication systems, high-quality video systems, imaging systems, and instrumentation systems. When higher accuracy but lower bandwidth is needed, oversampling techniques are typically employed. High OSRs are desirable since the requirements on the individual circuits are relaxed based on the OSR. However, increased bandwidth requirements necessitate techniques to reduce the OSR while still attaining good performance. The purpose of this paper is to investigate the effect of reducing the OSR of ΣΔ modulators and incremental data converters.

The most common oversampled data converter is the ΣΔ modulator which efficiently performs high-resolution data conversion and is typically reserved for high OSR applications where noise-shaping increases the signal-to-quantization noise ratio (SQNR). The difficulty at low OSRs is that noise-shaping is not as efficient because it increases the total noise power in the system thereby reducing the SQNR. Oversampled cascaded or MASH architectures provide an alternative as they are less sensitive to low OSRs than single-stage architectures [1] but are more sensitive to non-idealities in the circuit.

Nyquist-rate A/D converters can be used at low OSRs and they typically require some oversampling to reduce the requirements on the anti-aliasing filter at the input. However, they lose a significant advantage provided by noise-shaping since input-referred noise gets shaped by gain stages rather than the integrators present in noise-shaping converters. This fundamental advantage of ΣΔ modulators over pipeline A/D converters will be explored.

T. C. Caldwell (✉)
Analog Devices, University of Toronto, Toronto, Canada

H. Casier et al. (eds.), *Analog Circuit Design,*
DOI 10.1007/978-94-007-0391-9_8, © Springer Science+Business Media B.V. 2011

2 Background Information

2.1 Delta-Sigma Modulators

ΣΔ Modulators employ both oversampling and noise-shaping to improve the accuracy of a low-resolution (as low as 1-bit) internal A/D converter, or quantizer. With the feedback loop the noise in the quantizer has a different transfer function to the output than the signal. This allows the designer to choose a filter that will shape the noise and keep it small in the band of interest while also keeping the signal unattenuated in this frequency range. Only a small portion of the frequency band is kept through digital filtering leaving little noise within the band of interest, resulting in a high-resolution A/D converter at a reduced speed. The noise transfer function (NTF) and the signal transfer function (STF) characterize the ΣΔ modulator. The order and shape of the transfer functions, the OSR, and the internal A/D converter resolution determine the resolution of the ΣΔ modulator. A general single-stage modulator is shown in Fig. 1.

One significant improvement in ΣΔ modulators is the use of an input feed-forward branch [2]. An extra feed-forward path is added from the input to the summer in front of the quantizer. With this architecture, the NTF remains unchanged, but the STF is unity if $L_0(z) = -L_1(z)$. Also, signal content at the loop filter output is minimized. The delay-free path from the input through the A/D and D/A comes back and subtracts from the input. No signal content enters the loop filter, and all that is left is the error signal E_1. Therefore, the loop filter output is only a function of E_1. E_1 will always be somewhat correlated with the input, but for a higher-resolution internal A/D converter the input will be less correlated, and distortion introduced by the loop filter will be less signal dependent. This is advantageous as low-distortion OTAs get more difficult to design with increasingly smaller power supplies [3]. The general structure of the input feed-forward ΣΔ is shown in Fig. 2.

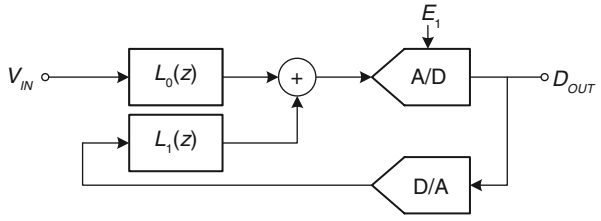


Fig. 1 General architecture of a ΣΔ modulator

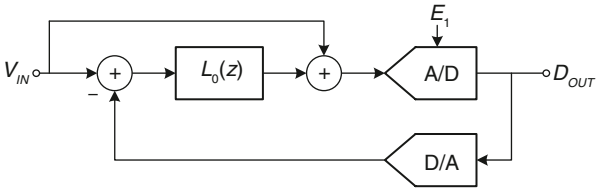


Fig. 2 Input feed-forward architecture for a general ΣΔ modulator

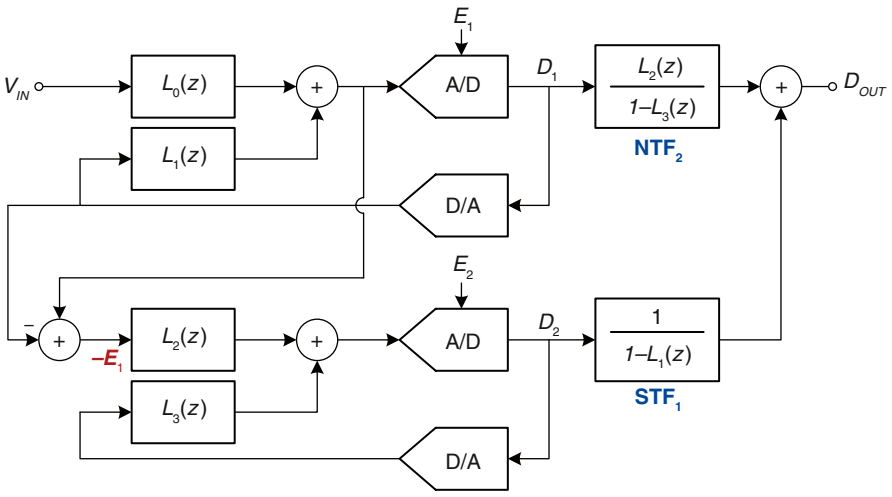


Fig. 3 General cascaded ΣΔ modulator

For stability reasons, it is difficult to design a high-order NTF and keep it stable as the filter coefficients vary, or the loop becomes nonlinear. There is also a trade-off between the NTF stability and how aggressively it shapes the noise (i.e., how much resolution it obtains). For these reasons high-resolution or low OSR ΣΔ modulators are often implemented with cascaded architectures. As the name implies, cascaded ΣΔ modulators cascade two or more single-stage ΣΔ modulators.

A general two-stage cascaded ΣΔ modulator is shown in Fig. 3. The advantage of cascaded modulators is that no individual modulator needs to be designed with a high-order filter; the total filter order can be spread out across many different stages so that each individual ΣΔ stage will be of lower order. The stability will be a function of the individual lower-order modulators rather than the total order of the modulator.

A digital filter is required to recombine the digital outputs of the individual ΣΔ modulators. It is designed to cancel the error introduced in the first stages, leaving only the error introduced in the last of the cascaded stages which is noise-shaped by the product of the NTF of each stage. This cancelation is dependent on matching between the digital filters and the analog filters within the individual ΣΔ modulators; this is one of the major limitations for high-resolution cascaded ΣΔ modulators.

As mentioned above, when the input feed-forward architecture is used the loop filter output contains no signal component. Depending on the quantizer resolution, this error signal might be considerably smaller than the input range of a ΣΔ modulator. When the error signal is passed to the subsequent stage, it may be possible to amplify it while staying within the allowable input range. The amplification factor will increase the overall resolution of the cascaded ΣΔ by the same amount.

2.2 Incremental Data Converters

Incremental A/D converters are best understood as a combination of ΣΔ modulators and dual-slope A/D converters. They act like dual-slope A/D converters mixed in time, but also have the benefit of utilizing higher-order loop filters like ΣΔ modulator.

The dual-slope (or integrating) A/D converter is useful for high-accuracy, high-linearity conversion with low offset and gain errors [4]. Shown in Fig. 4, the converter integrates the input signal for a fixed time and then subtracts a reference voltage for a counted number of clock periods until the output crosses zero. The final count at the zero crossing is the resulting digital output. For an N-bit A/D converter, 2^{N+1} cycles are required for one conversion. For high-resolution A/D converters the conversion time can severely limit the speed at which they operate.

The architecture of an incremental A/D converter is almost identical to that of a ΣΔ modulator except the integrators are reset after each conversion, the input is held for each conversion (ideally, but not always in practice), and the decimation filter

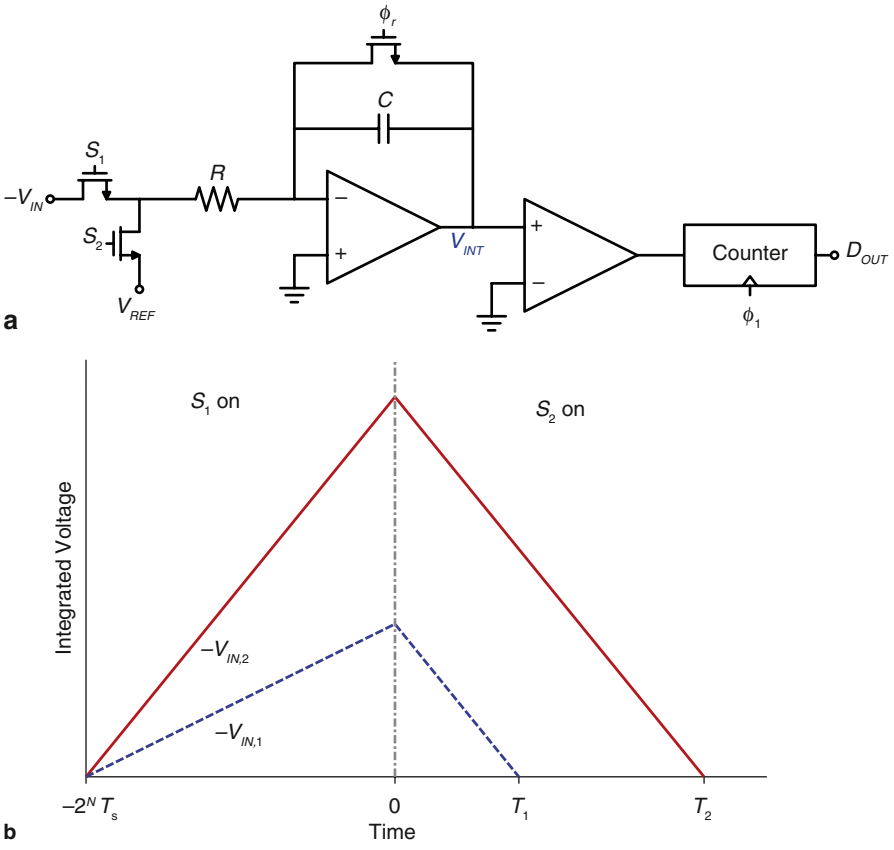


Fig. 4 Dual-slope A/D converter. **a** Architecture. **b** Time domain output

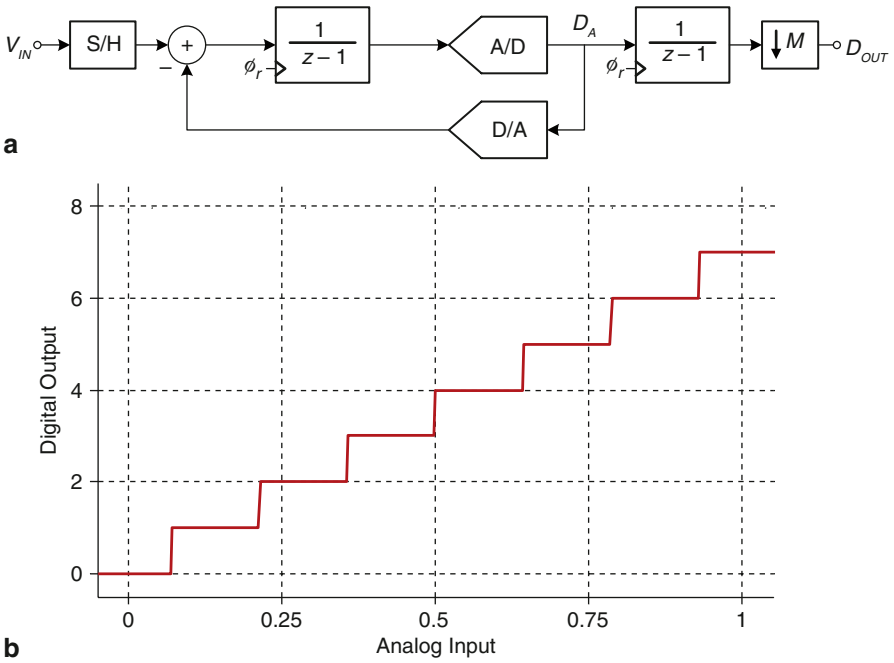


Fig. 5 Operation of a 1st-order incremental A/D converter with an OSR of 7. **a** Architecture. **b** Output vs. Input plot

is different. However, a 1st-order incremental A/D converter is better understood as operating like a dual-slope A/D converter since the input/output relationship is identical. Also, like a dual-slope A/D converter (and similar to Nyquist-rate A/D converters), input signals that fall between $f_s/2OSR$ and $f_s/2$ alias back into the signal band and are not suppressed by the digital decimation filter as they would be in a ΣΔ modulator.

A 1st-order incremental A/D converter is shown in Fig. 5. In contrast to a dual-slope A/D converter, the integration and subtraction of the reference signal are mixed in time. This is apparent if the 1-bit D/A converter output is either V_{REF} or 0 and the A/D threshold is V_{REF}. Assuming a positive unipolar input, the constant input is integrated until it is larger than V_{REF}. At this point, V_{REF} is subtracted from the input, and the counter is incremented by 1. This continues for $2^N - 1$ clock cycles to obtain a resolution of N bits (this is half as many as a dual-slope converter because a 2-phase clock is used). Once the conversion is performed (after OSR clock cycles), the integrator is reset and the next sample is converted. An added benefit to incremental A/D converters is the simplicity of the decimation filter. It can be as simple as a cascade of L accumulators for an Lth-order converter, although more complicated filters can be used [5–7].

An example of the converter output is shown in Fig. 5 for seven cycles, resulting in a 3-bit or 8-level output. While the incremental A/D converter is an oversampled

A/D converter, its characteristics are more similar to that of a Nyquist-rate A/D converter because input signals larger than $f_s/2OSR$ will alias back into the signal band unattenuated, assuming the use of an input sample-and-hold (S/H). In some cases the incremental A/D converter is considered a ΣΔ modulator operating in transient mode [8]. At high oversampling ratios, the longer an incremental A/D converter operates after the reset phase, the more similar its internal operation becomes to that of a ΣΔ modulator. However, at low OSRs (for example, an OSR of 3), it is most appropriate to think of an incremental A/D converter as operating distinctly from a ΣΔ modulator.

While the decimation filter of an incremental A/D converter can be as simple as a cascade of L accumulators for an Lth-order converter, more optimal filters exist. Reference [6] suggests the use of a dither signal in the analog loop filter along with a higher-order cascade of accumulators (L + 1) to improve the resolution. Reference [7] presents an optimal decimation filter that both minimizes the maximum error and the mean-squared error. These filters have increased digital complexity.

ΣΔ modulator techniques can be applied to incremental A/D converters. If the OSR of an incremental A/D converter is defined as the number of cycles in one conversion, then it is clear that an increased OSR will increase the resolution. Unlike dual-slope A/D converters, incremental A/D converters can utilize higher-order loop filters to further increase the resolution. Higher-order incremental A/D converters can be implemented in either a single-stage structure, or a cascaded or MASH architecture.

Single-stage architectures suffer from increased signal swings at the integrator outputs [9], but low-distortion input feed-forward architectures can be used to reduce these signal swings. At high OSRs, the valid input range for incremental A/D converters is similar to ΣΔ modulators which are governed by [1]

$$\max |V_{IN}| = (N + 1 - |h(n)|_1)/(N - 1) \tag{1}$$

for an N-level quantizer where $|h(n)|_1$ is the first norm of the NTF H(z).

Like ΣΔ modulators, the cascaded architecture is more stable for higher-order converters. As opposed to feeding the first integrator output into the subsequent stage, the first stage error can be fed into the following stage. This results in a smaller signal amplitude being fed to the subsequent stage since it has less signal component, and facilitates the use of interstage gains and multi-bit quantizers to increase the SQNR.

3 High-Order MASH ΣΔ Modulator

It is increasingly difficult to achieve high resolution in ΣΔ modulators at low OSRs. ΣΔ modulation is reliant on oversampled noise-shaping to increase the modulator's resolution, but noise-shaping increases the total quantization noise power in the

modulator, and the lower the OSR, the more significant this increased noise power becomes.

3.1 Comparison with a Nyquist-Rate A/D Converter

Since noise-shaping increases the total noise power in a ΣΔ modulator, at an OSR of 1 a Nyquist-rate A/D converter outperforms a ΣΔ modulator. However, as the OSR increases and less noise power ends up in the band of interest, the SQNR of the ΣΔ modulator improves at a faster rate than the 3 dB per octave improvement from the Nyquist-rate A/D converter. Eventually the ΣΔ modulator will outperform the Nyquist-rate A/D converter.

Figure 6 shows the total in-band integrated noise power (normalized) as a function of the OSR in a 1st-order and 8th-order ΣΔ modulator, as well as a Nyquist-rate A/D converter. The NTFs used in this comparison are $(1 - z^{-1})^L$ where L is the modulator order. Also it is assumed that the input can go as large as full-scale, although this would depend on the architecture and quantizer resolution. Also included in the comparison is a Nyquist-rate A/D converter where the normalization assumes the Nyquist-rate A/D converter has the same resolution as the total quantization noise in the ΣΔ modulator (i.e., both A/D converters have the same number of quantizer levels). There is a crossover point depending on the modulator order; for the 1st-order modulator this crossover occurs at an OSR of 1.66, and for an 8th-order modulator it occurs at an OSR of 2.48. This point defines the OSR where it is worth employing noise-shaping rather than simple oversampling for a given modulator order (assuming the same quantizer resolution can be used in both the ΣΔ modulator and the Nyquist-rate A/D converter).

The crossover point occurs at lower OSRs for lower-order ΣΔ modulators and the crossover point for various modulator orders is summarized in Table 1. Also included in the table is the relative reduction in noise power at an OSR of 3. Despite having a higher crossover OSR for the higher-order modulators, the higher-order

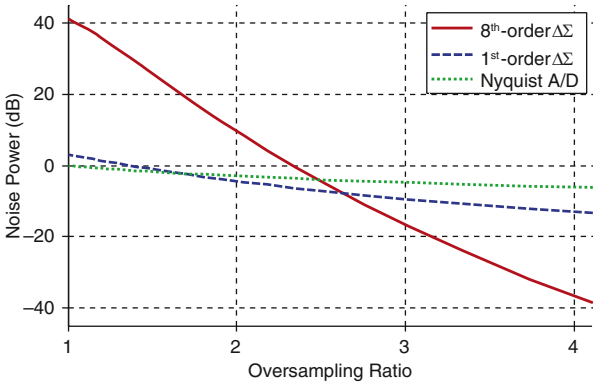


Fig. 6 Total in-band noise power vs. OSR

Table 1 OSR where ΣΔ and Nyquist-Rate A/D have equal noise power

NTF	Crossover OSR	Relative noise power at an OSR of 3 (dB)
1	1	−4.78
$(1 - z^{-1})$	1.66	−9.38
$(1 - z^{-1})^2$	1.94	−11.53
$(1 - z^{-1})^4$	2.23	−14.01
$(1 - z^{-1})^8$	2.48	−16.70
$(1 - z^{-1})^{12}$	2.60	−18.34

modulators still have lower noise powers at an OSR of 3 because the noise power decreases rapidly as the OSR increases. While it may appear that the reduction in noise power is not significant as the modulator order is increased, this table assumes that every architecture has the same quantization noise power. However, higher-order cascaded architectures are more amenable to higher resolution quantizers, resulting in smaller quantization noise power which further reduces the noise power in the modulator.

3.2 Single-Stage vs. Cascaded ΣΔ

In a ΣΔ modulator, improvements in SQNR come from three main factors: increased OSR, higher resolution in the quantizer, and higher modulator order. For a given OSR, and assuming a maximum quantizer resolution of about 4–5 bits (to preserve the simplicity of the design), the modulator order is the most influential degree of freedom for the designer, as well as its associated NTF. Since the SQNR is reduced with low OSRs, the modulator order must be increased for medium to high resolution ΣΔ modulators. Both single-stage and cascaded ΣΔ modulators are capable of realizing the same high-order NTFs, but cascaded ΣΔ modulators are more stable than single-stage architectures.

Increasing the modulator order in single-stage architectures typically requires increased resolution in the quantizer to maintain stability. This can require unreasonably large quantizers for high-order modulators, making single-stage architectures difficult for low OSR design. Figure 7 shows the peak SQNR for a 4th-order, 8th-order, and 12th-order single-stage modulator with an OSR of 3 and quantizer resolutions that vary from 3- to 1025-levels (approximately 10 bits) using the ΣΔ Toolbox [10]. For higher-order modulators (8th-order and 12th-order), a minimum quantizer resolution of 5-bits is necessary for an SQNR of at least 62 dB. The 4th-order modulator (and lower order modulators) would need quantizer resolutions of 65-levels or larger for a 10-bit modulator, which is prohibitively large since a flash A/D converter with such high resolution would take more power and require large comparators for well-matched devices, increasing the power consumption of the latter stages in the ΣΔ modulator.

A cascaded architecture can realize the same NTFs as single-stage architectures, but they are more stable since each individual stage can be of lower order

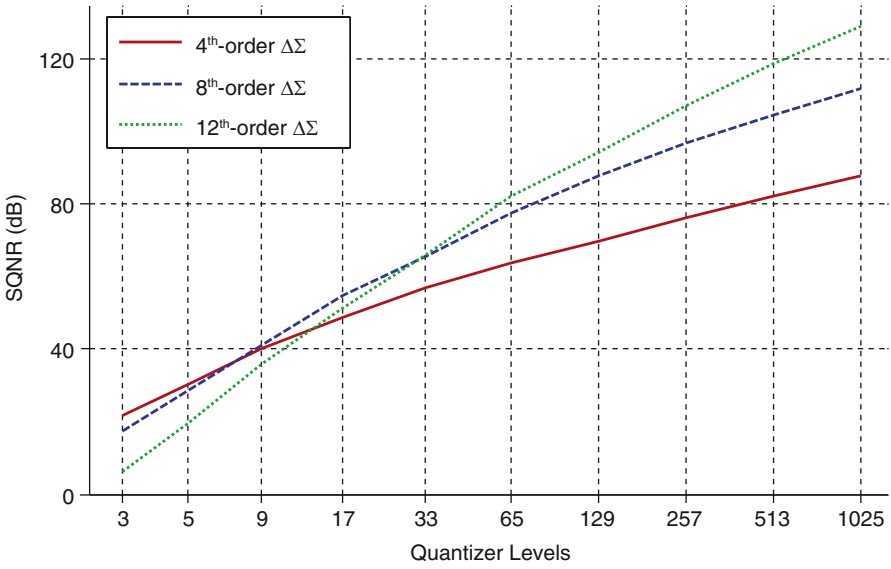


Fig. 7 SQNR vs. Quantizer levels

and stability of the entire modulator is a function of the individual stages' stability. This results in a larger full-scale input signal, as well as smaller quantizer resolutions in the individual stages which can still cascade to a total resolution larger than that of a single-stage architecture. However, the difficulty with MASH architectures is their reliance on matching between the analog and digital circuitry. The digital circuitry in a MASH ΣΔ must match the analog filters (integrators) for the appropriate noise cancelation to occur. This is increasingly difficult at lower OSRs since input-referred errors are more significant, but it is a familiar problem to Nyquist-rate pipeline A/D designers where the digital gains must match the analog gain circuits. The problem and solution are no different for a high-order MASH ΣΔ; either the OTA gain and capacitor matching must be sufficient, or calibration is required.

3.3 Sample Architecture

If a large number of 1st-order 3-level quantizer ΣΔ stages are cascaded, each individual stage is stable with a full-scale input. The quantizer resolution does not need to be increased to improve stability, and an interstage gain factor of 2 can be used between each stage to increase the resolution of the entire modulator. Each interstage gain of 2 increases the SQNR by 1 bit.

As an example, an 8th-order MASH ΣΔ modulator with an OSR of 3 is shown in Fig. 8. It cascades eight 1st-order ΣΔ stages with 3-level quantizers resulting in an

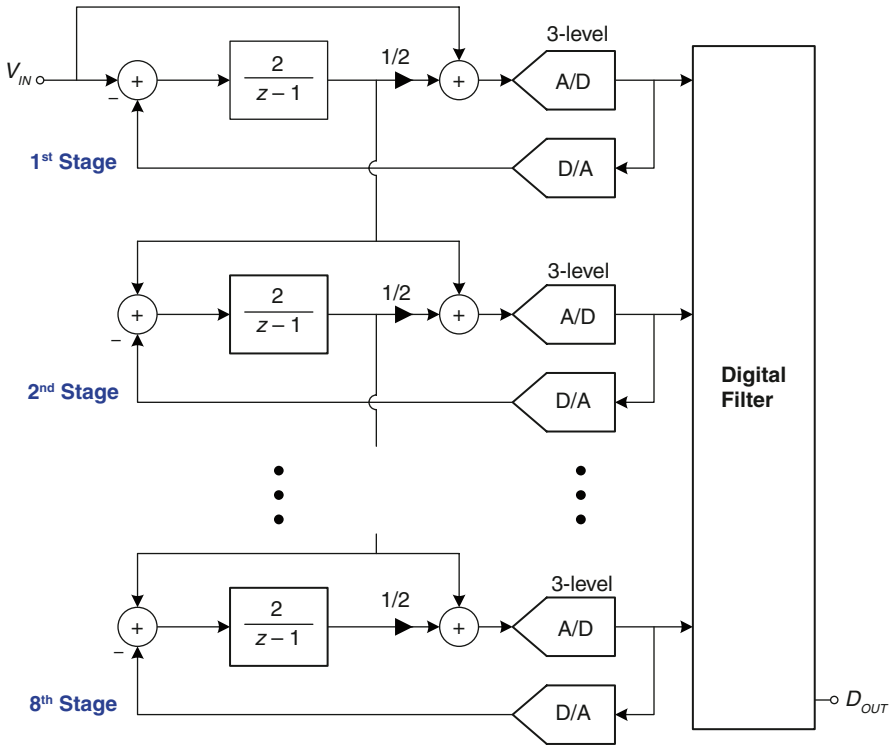


Fig. 8 8th-order MASH ΣΔ architecture

NTF of $(1 - z^{-1})^8$. Every stage is designed identically, making it relatively simple to implement for an 8th-order modulator. Each individual stage has an input feed-forward path to reduce the input-dependent signal at the integrator output. There is an interstage gain of 2 between each stage.

An 8th-order single-loop architecture could also obtain the same NTF of $(1 - z^{-1})^8$, but it has a few disadvantages. With a 257-level quantizer (which is equivalent to an 8-stage cascade of 3-level quantizers with interstage gains of 2) the modulator full-scale would be −14 dBFS, as opposed to 0 dBFS for the MASH architecture. Not only does this reduce the resolution by 14 dB, but it requires the unreasonably large 257-level quantizer. There are better ways to implement an 8th-order single-loop transfer function (for example, optimized zeros, optimal pole placement), but if the NTFs are kept the same it is clear that a MASH architecture is far more suitable to low OSR design.

A sample output spectrum of the 8th-order MASH ΣΔ is shown in Fig. 9. The architecture achieves a peak SQNR of 66 dB. The NTF is not optimized for the OSR because only 1st-order stages are used; 2nd-order stages would be necessary to create resonator structures that optimize the NTF zeros.

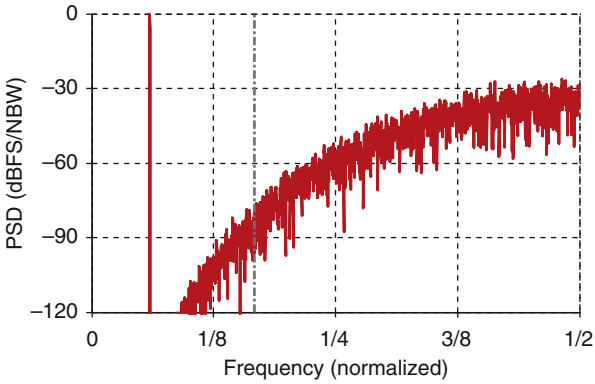


Fig. 9 Output spectrum of an 8th-order MASH ΣΔ (NBW = 3.7×10^{-4})

3.4 Power Efficiency

The proposed architecture is more power efficient than a comparable Nyquist-rate A/D converter, specifically a 1.5 bit/stage pipeline A/D converter. These two architectures are almost identical in that each requires one OTA per stage and one 3-level quantizer per stage. Also, assuming a power efficient design, they should both be thermal noise limited.

A large fraction of the power for switched-capacitor thermal noise limited A/D converters is from the first stage since it should be equally power efficient irrespective of the targeted speed or resolution of the switched-capacitor circuit (assuming similar topologies, to a first-order approximation). The first stage of any A/D converter with a given resolution and bandwidth requires the same amount of analog power, whether it is oversampled or not. Since the sampling capacitor of an oversampled modulator is OSR times less than that of a Nyquist-rate pipeline A/D converter, the power required by the OTA of the oversampled modulator is OSR times less, but since the sampling frequency is OSR times more, the resulting power is equivalent.

When the power of subsequent stages is included, the more power efficient architectures are those with a lower input-referred noise from later stages. Generally, ΣΔ modulators will have lower input-referred noise from later stages when compared to pipeline A/D converters since the noise from subsequent stages is input-referred through integrators rather than gain stages. Figure 10 shows the input-referred noise through a gain stage and an integrating stage. At an OSR of 3 the two gains are the same, but the total input-referred noise is the area under the curves in the signal band and is less in the integrator/ΣΔ case. However, other input-referred parameters such as linearity, settling accuracy, etc. are no different at an OSR of 3 because they must be designed for the worst case at the signal band edge, resulting in the same requirements for both a pipeline A/D converter and a ΣΔ modulator.

As an example, if an 8th-order input feed-forward MASH ΣΔ with an OSR of 3 is compared against a 10-stage, 1.5-bit/stage pipeline A/D converter operating at the same sampling frequency with the same OSR of 3, then both A/D converters will

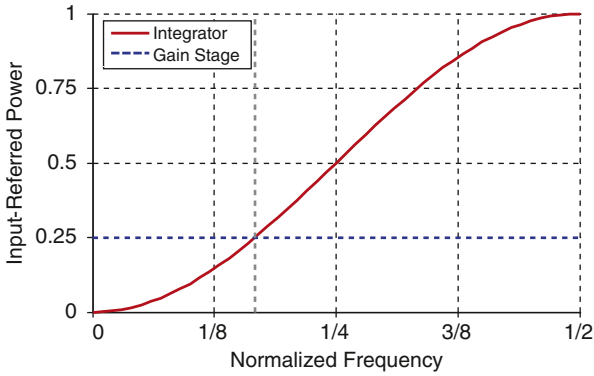


Fig. 10 Input-referred noise comparison between a gain and integrator stage

have 66–67 dB of SQN while the required analog circuitry for these architectures is similar (although the pipeline A/D converter has two extra stages at the end). It is assumed that the pipeline A/D converter can do without the input S/H, a fair assumption given that the input feed-forward MASH ΣΔ has no S/H and should be similarly reliant on the matching of both input paths. If the design is such that the second and third stages are 2 times smaller than the first stage, and the fifth to last stages are eight times smaller, then for the same input-referred noise the pipeline A/D converter consumes 60% more power. If the first stage is normalized to a size of 1, the total size of the pipeline A/D converter is 2.875 while the total size of the MASH ΣΔ modulator is 2.625. With the first stage input-referred noise normalized to 1 (with no oversampling), the total input-referred noise from the 10-stage pipeline is 0.597 while the input-referred noise from the 8-stage MASH ΣΔ is 0.408. The input-referred noise of the pipeline A/D converter is 46% more, and taking into account the extra 10% power for the two extra stages, the total additional power of the pipeline A/D converter is 60%.

The assumption that the first stage power will be the same assumes a similar clocking scheme. However, the MASH ΣΔ architecture has the flexibility to use full-period delaying integrators with each stage clocked on ϕ_1. On the contrary, pipeline A/D converters are constrained to using half-delaying gain stages since a half clock cycle is needed to reset the previous value on the gain stage (i.e., odd stages of the pipeline sample on ϕ_1, even stages of the pipeline sample on ϕ_2). Half-delaying gain stages are power inefficient when the subsequent stage has a similarly sized capacitor because the subsequent stage loads the current stage while it is in the amplifying/integrating phase, and this the phase typically limits the OTA bandwidth since power efficient designs have a switch resistance that is smaller than $1/G_m$ of the OTA. To charge the subsequent stage's sampling capacitance, extra current must be drawn from the amplifier, resulting in a larger amplifier. Also, a larger amplifier causes a reduced feedback factor β, further increasing the amplifier size for a given settling time constant.

Comparing the same 1.5 bit/stage pipeline A/D converter against the architecturally similar 8th-order MASH ΣΔ, the power savings from full-delaying gain stages instead of half-delaying gain stages can be seen. As shown in Fig. 11, during the am-

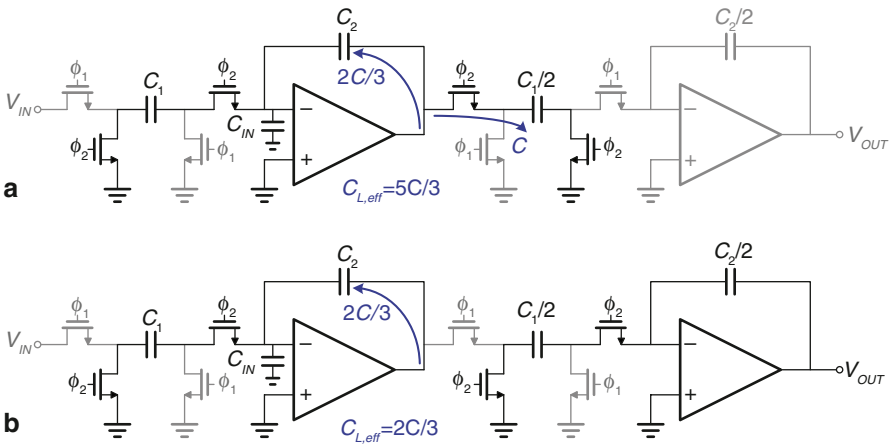


Fig. 11 Comparison of non-delaying and delaying stages. **a** Non-delaying stage (for a pipeline A/D converter). **b** Delaying stage (for a ∑Δ modulator)

plification/integrating phase the 3-dB frequency is $\omega_{3\ dB} = \beta G_m/C_{L,eff}$ where $\beta = C_2/(C_1 + C_2 + C_{IN})$ and for the half-delaying pipeline gain stage

$$C_{L,eff} = C_1/2 + C_2(C_1 + C_{IN})/(C_1 + C_2 + C_{IN}) \tag{2}$$

while for the full-delaying ΣΔ integrating stage

$$C_{L,eff} = C_2(C_1 + C_{IN})/(C_1 + C_2 + C_{IN}). \tag{3}$$

With $C_1 = 2C_2$ (for a 1.5-bit/stage design, or an integrator with a gain of 2), the effective load capacitance $C_{L,eff}$ and OTA power of the delaying stage is 2.5 times less than that of the non-delaying stage for the same $\omega_{3\ dB}$ assuming $C_{IN} = 0$. For increasing values of C_{IN} the disparity becomes larger because the non-delaying stage has a larger OTA than the delaying stage, resulting in larger C_{IN}, smaller β, and decreased OTA efficiency.

It may be considered unrealistic to assume that the ΣΔ stage has no load capacitance on the amplifying phase. However, even if a capacitance equal to 20% of the feedback capacitance C_2 is added as a parasitic capacitance, it still consumes less than half the power of the pipeline OTA for the same $\omega_{3\ dB}$.

A couple architectural improvements can be used to improve the power efficiency of a pipeline A/D converter. A pipeline gain stage has improved power efficiency with a precision gain stage since it has a reduced feedback factor and load capacitance in the amplification phase [11]. This is shown in Fig. 12 where $\beta = C_2/(2C_2 + C_{IN})$ and

$$C_{L,eff} = C_2 + C_2(C_2 + C_{IN})/(2C_2 + C_{IN}) \tag{4}$$

The resulting power is only 50% more with the same $\omega_{3\ dB}$ when compared to the delaying integrators of the ΣΔ modulator.

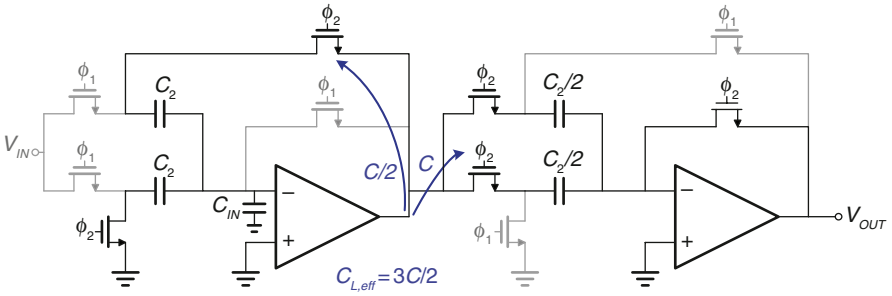


Fig. 12 Precision delaying gain stage

Other alternatives include powering down the OTA in the sampling phase using the switched-opamp technique [12], or sharing it with a parallel structure [13]. A further improvement uses the switched-opamp technique while reducing the amplifier loading by using the feedback capacitor of the current stage as the sampling capacitor for the subsequent stage [14]. With these techniques roughly half the power is required for the same $\omega_{3\ dB}$ (except for [14] which has the potential to make further improvements). These techniques are necessary to keep the pipeline A/D converter's power efficiency similar to that which is inherent in the MASH ΣΔ architecture.

4 Incremental Data Converters

4.1 Comparison of Incremental and ΣΔ Converters

At low OSRs, incremental data converters have higher SQNRs than ΣΔ modulators. The reason comes from the fact that the two converters operate on different principles, especially at low OSRs where the resetting nature of incremental data converters is more significant. A single-stage architecture has

$$1 + \alpha(N-1)(M+L-1)!/L!(M-1)! \tag{5}$$

output levels while a cascaded architecture has

$$1 + \alpha(N-1)^{L}(M+L-1)!/L!(M-1)! \tag{6}$$

output levels for an N-level quantizer, Lth-order converter, an OSR of M, and stable input range of α. The contrasting feature of incremental A/D converters when compared to ΣΔ modulators is that for any non-zero value of L, M and N, the number of output levels will be one or more. Unlike ΣΔ modulators at low OSRs where noise-shaping increases the total quantization noise power of the system, incremental A/D

Table 2 Comparison of ΣΔ and incremental A/D converters at low OSRs

OSR	Incremental		ΣΔ	
	MASH (dB)	Single-stage (dB)	MASH (dB)	Single-stage (dB)
2	26	24	17	11
4	35	33	31	25
8	46	44	46	40
16	57	54	59	56

converters will always have a minimum resolution equal to the quantizer resolution, even at an OSR of 1.

Table 2 compares simulated SQNRs of a ΣΔ modulator and an incremental A/D converter at an OSR of 2, 4, 8 and 16 using an NTF of $(1 - z^{-1})^2$ with a 2nd-order MASH architecture (with 3-level quantizers and an interstage gain of 2) as well as a 2nd-order single-stage architecture (with a 5-level quantizer). At a low OSR of 2 and 4, the incremental A/D converter has a larger SQNR, while at an OSR of 8, the results are similar. At higher OSRs of 16 and above, the ΣΔ modulator has a larger SQNR. These results motivate the use of incremental A/D converters instead of ΣΔ modulators at lower OSRs.

As another example, Fig. 13 shows an SQNR comparison at increasing OSRs for an 8th-order cascaded ΣΔ modulator, an 8th-order cascaded incremental A/D converter, and an 8-stage pipeline A/D converter. Each stage has a 3-level quantizer so that the three converters are almost identical architecturally. As long as the OSR is less than 5.3, the incremental A/D converter has a higher SQNR than the other two architectures. Beyond an OSR of 5.3, the noise-shaping in a ΣΔ modulator outperforms incremental A/D conversion.

Incremental A/D converters and ΣΔ modulators operate on different principles because the loop filter is reset and the input is held in an incremental converter. Since the incremental A/D converter has less noise than a ΣΔ modulator at an OSR of 1 due to the increased noise power from noise-shaping, it should not be surprising that the incremental A/D converter outperforms the ΣΔ modulator up until the OSR where noise-shaping begins to improve the ΣΔ modulator's resolution.

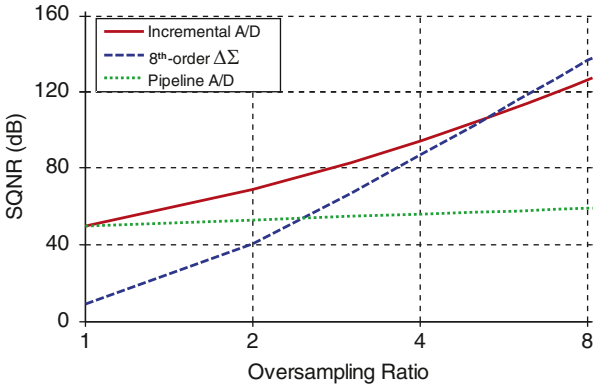


Fig. 13 Simulated SQNR vs. OSR for three different architectures

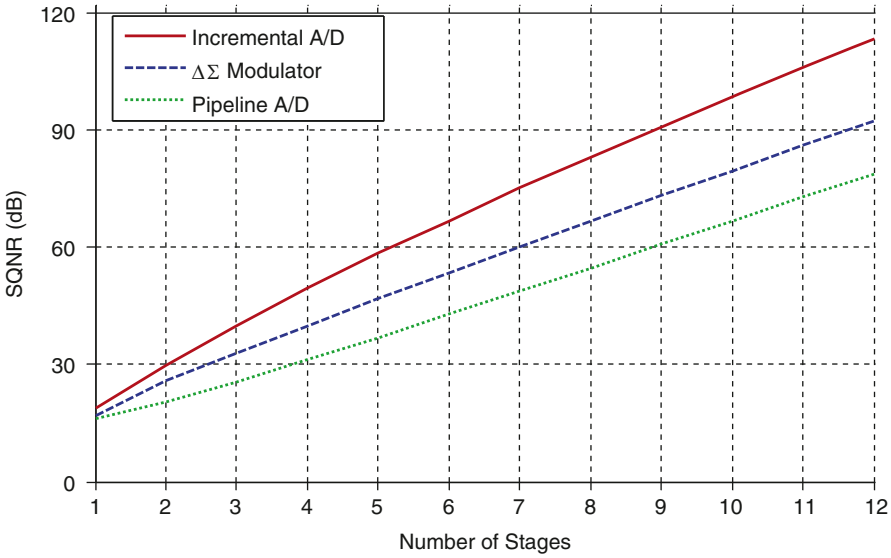


Fig. 14 Simulated SQNR vs. Number of Stages at an OSR of 3

Another reason for the increased resolution of an incremental A/D converter at low OSRs is the larger allowable signal amplitudes. Since an incremental A/D converter resets after OSR clock cycles, the memory of previous conversions is lost, and thus a sustained large signal has only OSR clock cycles to accumulate in the integrator to overload the quantizer. Therefore, while a ΣΔ limits the signal amplitude at low OSRs as it would at high OSRs, an incremental A/D converter at low OSRs allows larger signal amplitudes than at high OSRs since there are fewer cycles for the signal to accumulate. This can contribute a few extra dB of resolution as the OSR is lowered.

At an OSR of 3 (with 3-level quantizers), these three architectures can also be compared while varying the number of (1st-order) stages. The results are shown in Fig. 14. At this OSR the incremental A/D converter outperforms the ΣΔ modulator and pipeline A/D converter.

4.2 Pipeline Equivalency

As pointed out in the previous section, even with an OSR of 1, the incremental A/D converter still resolves the input signal with the internal quantizer. If a cascaded incremental A/D converter is used, Eq. 6 predicts $1 + \alpha(N - 1)^L$ output levels for an Lth-order converter with N-level quantizers. This is identical to the resolution of an L-stage pipeline A/D converter with N-level internal quantizers. In fact, when

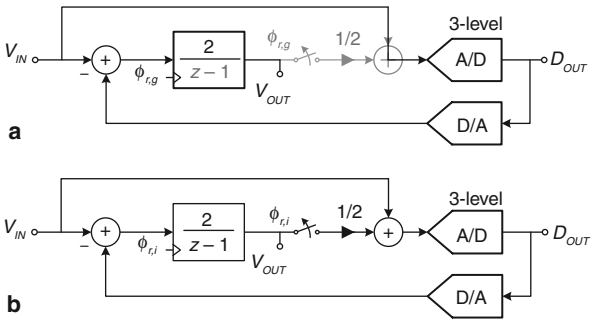


Fig. 15 Similarity of a pipeline stage and an incremental stage. **a** Resetting gain stage. **b** Resetting integrating stage

the incremental A/D converter has an OSR of 1, it is effectively a pipeline A/D converter.

More specifically, an input-feedfoward cascaded incremental A/D converter can be thought of as a pipeline A/D converter where the OSR determines how frequently resetting is performed. Architecturally the two are almost identical; the main difference lies in designing a gain stage or an integrating stage. A single stage of both architectures is shown Fig. 15 and they are almost identical. The incremental A/D converter stage uses a resetting integrator while the pipeline A/D converter uses a gain stage which is effectively an integrator that resets on every clock cycle $\phi_{r,g}$. Also, the addition at the quantizer input in the incremental converter occurs on all clock cycles except for the resetting phase $\phi_{r,i}$. This is also true for the pipeline converter stage, but since it resets on every clock cycle, this addition is never performed.

At the circuit level, a resetting gain stage and a resetting integrating stage are shown in Fig. 16. The gain stage clock $\phi_{r,g}$ resets on ϕ_1, while the integrator clock $\phi_{r,i}$ resets on ϕ_1, but only every Mth clock cycle (for an OSR of M). Aside from a couple switches, the only difference lies in the resetting sequence. With the same C_1 and C_2 (a reasonable assumption since C_1 determines the thermal noise, and C_2 controls the gain), both OTAs would be designed almost identically.

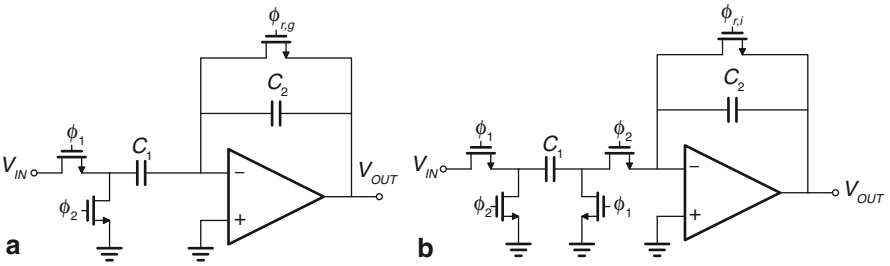


Fig. 16 Circuit-level similarity of a pipeline stage and an incremental stage. **a** Resetting gain stage. **b** Resetting integrating stage

4.3 Removing the Input S/H

An incremental A/D converter ideally requires an accurate S/H circuit on the input. The design of this block can be difficult, especially when trying to achieve better than 10-bit performance. A S/H circuit also costs extra power since an extra circuit is needed, and this circuit contributes more noise to the entire A/D converter, requiring increased power in the rest of the converter to lower the total noise. Removing the S/H results in a modified STF which causes some high-frequency attenuation. This can be understood by analyzing a 1st-order converter and extending the result to higher-order converters.

In a single-bit 1st-order incremental A/D converter it will be shown that the output of one conversion will always be the same as long as the average of the input for that conversion is constant (assuming no quantizer overload). For an input V_{IN}, assuming an input feed-forward architecture, the input to the quantizer after the first M clock cycles will be

$$V_Q[M] = \sum_{i=1}^{M} V_{IN}[i] - \sum_{i=1}^{M-1} V_{REF} D_1[i] \tag{7}$$

If it is assumed that the converter is operating within the converter input range where the quantizer is not overloaded, then $-2V_{REF} < V_Q[M] < 2V_{REF}$. This also means that if the last sample $D_1[M]$ is included in the inequality,

$$-V_{REF} < V_Q[M] - V_{REF} D_1[M] < V_{REF} \tag{8}$$

which is equivalent to

$$-V_{REF} < \sum_{i=1}^{M} V_{IN}[i] - \sum_{i=1}^{M} V_{REF} D_1[i] < V_{REF}. \tag{9}$$

For a given sum of inputs throughout the M cycles $\sum_{i=1}^{M} V_{IN}[i]$, there is a unique sum of digital outputs $D_{OUT} = \sum_{i=1}^{M} D_1[i]$ that will keep Eq. 9 bounded within V_{REF}. As long as $\sum_{i=1}^{M} V_{IN}[i]$ is constant, D_{OUT} will be constant. Since D_{OUT} is simply the final digital output of the incremental A/D converter after going through the accumulating decimation filter, this will be a unique digital output as long as the sum of the inputs $V_{IN}[i]$ is constant (which is equivalent to keeping the average of the input samples constant). If the input of the 1st-order incremental A/D converter is averaged and then passed through a S/H, this system will be identical to one where a moving input enters the system with no S/H since the output is only a function of the sum. So the incremental A/D converter with a moving input can be modeled as a typical incremental A/D converter where the S/H is preceded by an averaging filter G(z), as shown in Fig. 17. This provides a direct way to analyze the effect of the moving input on an incremental A/D converter using the filter G(z).

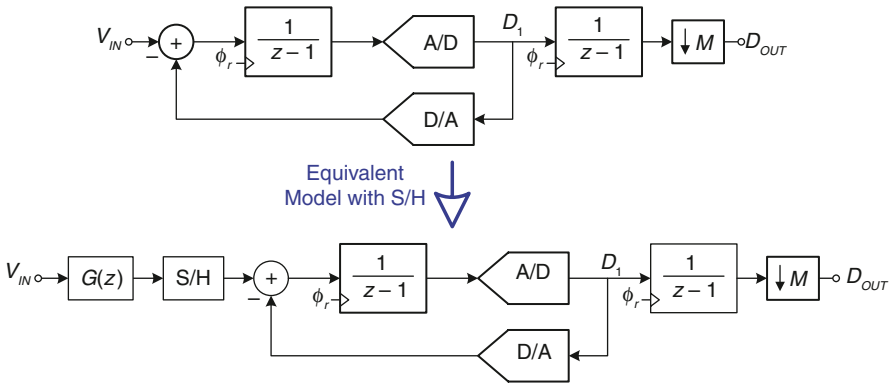


Fig. 17 Incremental converter with no S/H, and its equivalent model with a S/H

The filter G(z), while not explicitly present in the incremental A/D converter, effectively modifies the STF when the input S/H is removed. Depending on the modulator order, G(z) will have a different shape. The filter is of the form

$$G(z) = \left(1 + z^{-1} + z^{-2} + \cdots + z^{-(M-1)}\right)/M \tag{10}$$

for a 1st-order incremental with an OSR of M. Figure 18 shows G(z) with an OSR of 3. Since the input signal is limited to $f_s/2OSR$, the attenuation will be no larger than 3.52 dB at the signal band edge, represented by the vertical dotted line in Fig. 18. The filter is a digital sinc filter similar to what is seen in a dual-slope A/D converter (which inherently has no S/H). While input signals between $f_s/2OSR$ and $f_s/2$ will still alias back into the signal band, they will be attenuated according to the STF shown in Fig. 18.

The discussion can be extended to find STFs of higher-order converters with removed S/H blocks. For these architectures the STF is a weighted sum of inputs $V_{IN}[i]$ that are held constant, resulting in a more complicated filter G(z) that is a

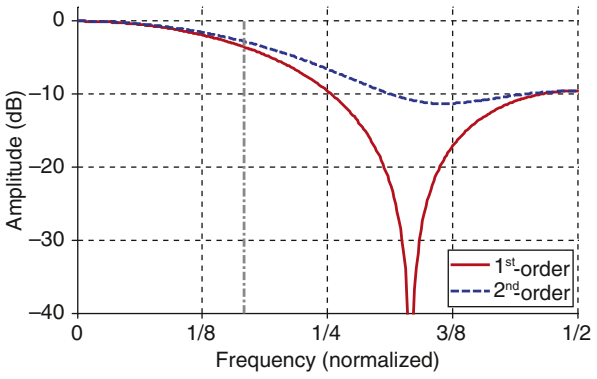


Fig. 18 G(z) for a 1st- and 2nd-order incremental converter with an OSR of 3

weighted average of the inputs. The resulting equivalent filter G(z) for a 2nd-order incremental A/D converter with a moving input is

$$G(z) = 2(1 + 2z^{-1} + 3z^{-2} + \cdots + Mz^{-(M-1)})/(M^2 + M) \tag{11}$$

The STF of this incremental A/D converter for an OSR of 3 is also shown in Fig. 18. The attenuation is 2.78 dB at the signal band edge. The attenuation at the signal band edge is less and will continue to reduce for higher-order incremental A/D converters with the input S/H removed.

To maintain a unity-gain in the signal band an additional digital filter can be added which is equal to the inverse of G(z) throughout the signal band. This increases the quantization or thermal noise towards the signal band edge but the total increase in noise is small since the majority of the noise is at lower frequencies and does not get amplified. For example, with the 1st-order converter of Fig. 18 at an OSR of 3 the noise increases by only 1.05 dB, and this will be less for higher-order converters.

4.4 Sample Architecture

As has been discussed, high-order cascaded architectures are better suited to low OSR incremental A/D designs. A sample architecture is shown in Fig. 19 and is

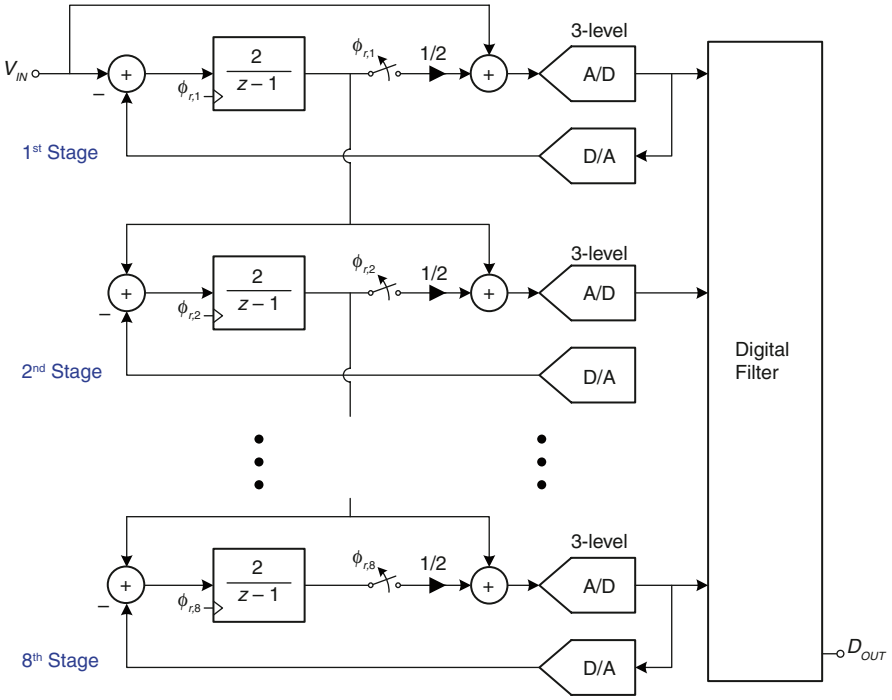


Fig. 19 Sample incremental A/D architecture

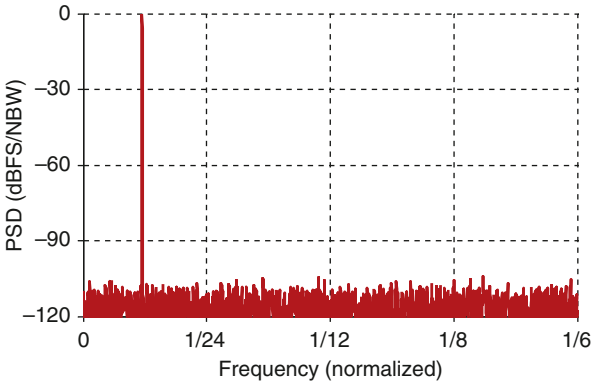


Fig. 20 Output spectrum for an 8th-order incremental A/D architecture

almost identical to the ΣΔ architecture of Fig. 8. It is an 8th-order cascade of 1st-order stages with 3-level quantizers and an OSR of 3. With the input S/H removed, the STF of the proposed architecture has a maximum attenuation of 0.97 dB at the signal band edge. This is relatively small and would likely be considered a worthwhile trade-off since the high power S/H is no longer needed. A sample plot of the downsampled spectrum is shown in Fig. 20. It has a peak SQNR of 83.5 dB, much larger than the 66 dB SQNR of the ΣΔ modulator.

5 Conclusions

In this paper it was shown that an 8th-order MASH ΣΔ modulator with an OSR of 3 can be used for medium resolution data conversion using a cascade of 1st-order ΣΔ stages. They were also shown to be a viable alternative to pipeline A/D converters. Also, theory for designing incremental A/D converters at low OSRs was presented. It was shown that an incremental A/D converter has improved performance over a ΣΔ modulator at very low OSRs, and is equivalent to a pipeline A/D converter when oversampled by unity.

References

1. R. Schreier, G.C. Temes, *Understanding Delta-Sigma Data Converters*. (Wiley, New Jersey, 2005)
2. J. Silva, U. Moon, J. Steensgaard, G.C. Temes, Wideband low-distortion delta-sigma ADC topology. IEE Electron. Lett. **37**, 737, 738 (2001)
3. J. Silva, U. Moon, G.C. Temes, Low-distortion delta-sigma topologies for MASH architectures, in *Proc. IEEE International Symposium on Circuits and Systems*, May 2004, pp. I1144–I1147
4. D.A. Johns, K. Martin, *Analog Integrated Circuit Design*. (Wiley, Toronto, 1997)
5. J. Steensgaard, Z. Zhang, W. Yu, A. Sarhegyi, L. Lucchese, D. Kim, G.C. Temes, Noise-power optimization of incremental data converters. IEEE Trans. Circuits Syst. I. **55**, 1289–1296 (2008)

6. J. Markus, J. Silva, G.C. Temes, Theory and applications of incremental ΣΔ converters. IEEE Trans. Circuits Syst. I. **51**, 678–690 (2004)
7. S. Kavusi, H. Kakavand, A.E. Gamal, On incremental sigma-delta modulation with optimal filtering. IEEE Trans. Circuits Syst. I. **53**, 1004–1015 (2006)
8. V. Quiquempoix, P. Deval, A. Barreto, G. Bellini, J. Collings, J. Markus, J. Silva, G.C. Temes, A low-power 22-bit incremental ADC. IEEE J. Solid-State Circuits, **41**, 1562–1571 (2006)
9. J. Robert, P. Deval, A second-order high-resolution incremental A/D converter with offset and charge injection compensation. IEEE J. Solid-State Circuits, **23**, 736–741 (1988)
10. R. Schreier, The Delta-Sigma Toolbox for Matlab (2009), http://www.mathworks.com/matlabcentral/fileexchange/19-delta-sigma-toolbox (The toolbox is generated in 2000 and updated in 2009)
11. B. Razavi, *Design of Analog CMOS Integrated Circuits*. (McGraw-Hill, New York, 2001)
12. J. Crols, M. Steyaert, Switched-opamp: an approach to realize full CMOS switched-capacitor circuits at very low power supply voltages. IEEE J. Solid-State Circuits, **29**, 936–942 (1994)
13. K. Nagaraj, H.S. Fetterman, J. Anidjar, S.H. Lewis, R.G. Renniger, A 250-mW, 8-b, 52-Msamples/s parallel-pipelined A/D converter with reduced number of amplifiers. IEEE J. Solid-State Circuits, **32**, 312–320 (1997)
14. N. Sokal, A. Sokal, Class E—A new class of high-efficiency tuned single-ended switching power amplifiers. IEEE JSSC. **10**(3), 168–176 (1975)

Comparator-Based Switched-Capacitor Delta-Sigma A/D Converters

Koen Cornelissens and Michiel Steyaert

1 Introduction

One of the major difficulties for analog circuit design in nanometer CMOS technologies is the design of OTAs. Nanoscale CMOS technologies have a low supply voltage and the transistors have a small output resistance. This makes it difficult to achieve high gain. For this purpose, cascodes are often used. However, with a low supply voltage, this limits the output swing. An alternative is to cascade multiple stages. However, this significantly increases power consumption as large compensation capacitors are required for stability in feedback configurations.

A ΔΣ A/D converter has a noise-shaping filter to generate a different transfer function for the signal and for the quantization noise. This allows to push the quantization noise out of the signal band and to increase the SNDR. This filter is often implemented using switched-capacitor techniques. In such a switched-capacitor noise-shaping filter, OTAs are required to create a virtual ground node, what allows transferring charge from one capacitor to another.

For such OTAs a large output swing is a necessity. This follows from thermal noise considerations. If the signal swing is halved, the signal power divides by four. Consequently, for the same SNR, the noise power must also be reduced with a factor of four. This requires making all capacitors four times larger [1]. Thus, for the same speed, the OTA needs four times more current. Therefore, in nanoscale CMOS technologies, which are already limited by a low supply voltage, it is mandatory to have signal swings as large as possible.

Specific methods to design OTAs in nanoscale CMOS technologies have been proposed. Current bleeding [2] and class A/B output stages [3] allow maintaining reasonable power consumption. The use of body-driven transistors [4] allows lower supply voltages. However, this technique suffers from higher noise levels and lower transconductances. In [5], the OTA is replaced by inverters. The large offset is cancelled using auto-zeroing. However, to obtain sufficient gain, biasing in weak inversion is necessary, limiting the bandwidth.

K. Cornelissens (✉)
ESAT-MICAS, K.U. Leuven, Kasteelpark Arenberg 10, 3001 Leuven, Belgium

H. Casier et al. (eds.), *Analog Circuit Design,*
DOI 10.1007/978-94-007-0391-9_9, © Springer Science+Business Media B.V. 2011

This paper discusses a technique to construct switched-capacitor filters without OTAs. The OTA is replaced by a comparator and a current source. As the gain and stability requirements are completely different, such a comparator-based switched-capacitor (CBSC) is well suited for design in nanoscale CMOS technologies. Whereas this has been previously done for pipeline A/D converters [6–9], this paper focuses on the use of CBSC circuits for ΔΣ A/D converters.

Section 2 discusses the implementation of a CBSC switched-capacitor integrator. The principle is explained and compared with an OTA-based implementation. In Sect. 3, the requirements for the comparator used in a CBSC integrator are derived. Section 4 discusses the requirements and implementation of the current sources needed in the integrators. In Sect. 5, a noise-shaping filter is constructed using CBSC integrators. An implementation of a fourth-order CBSC ΔΣ A/D converter in a 1 V, 90 nm CMOS technology, together with the measurement results, is presented in Sect. 6. Finally, conclusions are drawn in Sect. 7.

2 Switched-Capacitor Integrator

In Fig. 1, a standard OTA-based integrator is shown. During the sampling phase (Φ_1), the input voltage is sampled on C_s, while the integrated value is held on C_i. At the beginning of the integration phase (Φ_2), charge redistribution between C_s and C_i takes place. The OTA will react by delivering current until V_a is back at the virtual ground voltage. At that moment, all charge that was initially on C_s is transferred to C_i.

The gain of the OTA determines the accuracy of this charge transfer. The settling speed is determined by the ratio of the transconductance of the OTA (g_m) to the sampling capacitor (C_s). For the design of ΔΣ A/D converters, the required gain and settling speed are often derived using behavioral models [1].

In an OTA-based implementation, the OTA will always work to bring both input voltages at the same level. However, for a switched-capacitor system this is not

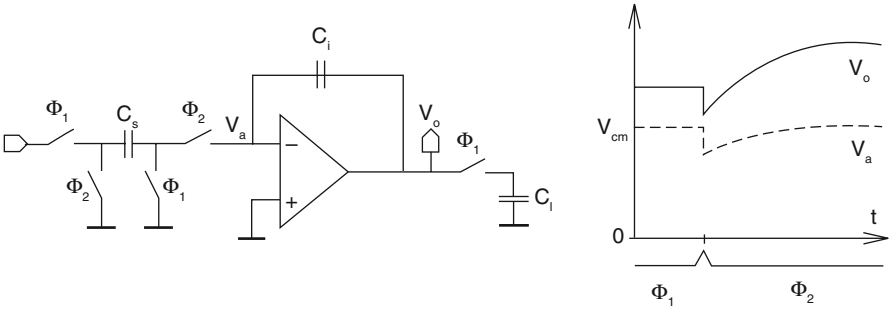


Fig. 1 An OTA-based integrator

necessary. In order to have the correct transfer function, the OTA inputs only need to be at the same voltage at the end of the clock phase. The settling behavior during the clock phase has no importance. This is exploited in a CBSC system. With a current source, the output voltage is swept, until a comparator detects a threshold crossing of the input. Then, the current source is shut down so that the node voltages cannot change anymore.

Figure 2 shows three different possible implementations for a CBSC integrator. Although they all implement the same transfer function, the sensitivity to non-idealities is different. Compared to the OTA-based implementation, one switch less is needed. In an OTA-based implementation, that switch is necessary to prevent the OTA from discharging C_i during Φ_1. For the CBSC implementations shown in

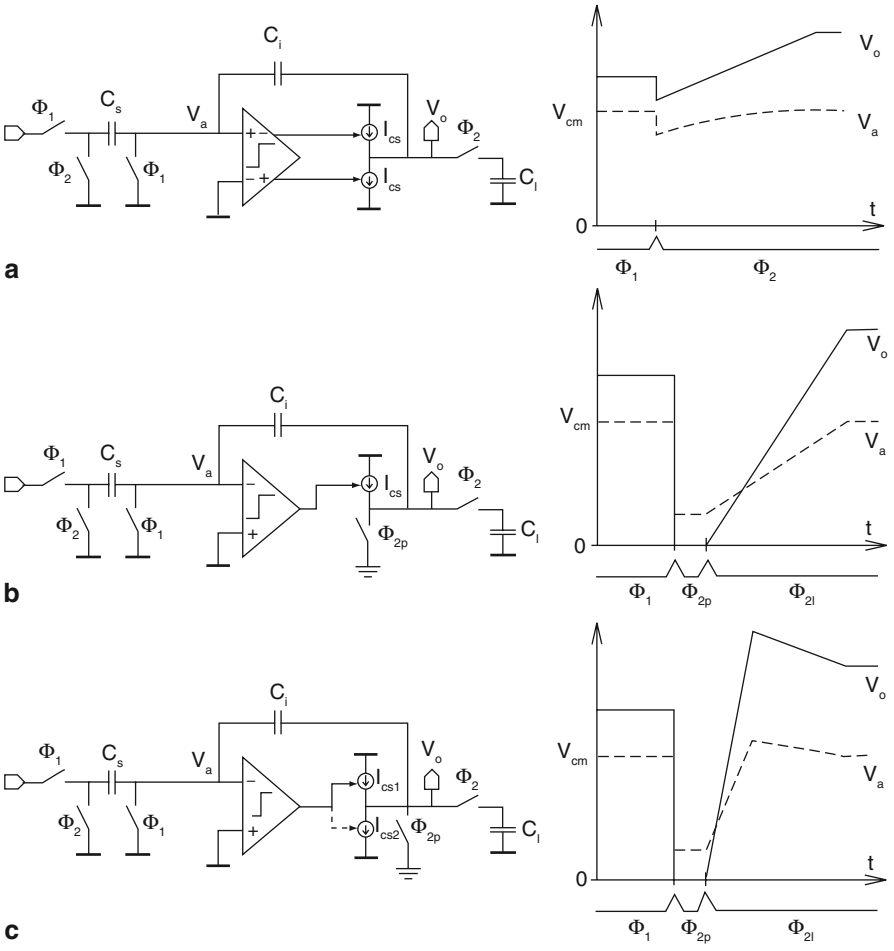


Fig. 2 Different possible implementations of a CBSC integrator. **a** Two directional current source. **b** Precharge and single current source. **c** Precharge and double current source

Fig. 2, the current sources are shut down during Φ_1. Hence, the charge on C_i cannot change during Φ_1.

The implementation shown in Fig. 2a uses two current sources. At the beginning of Φ_2, charge redistribution between C_s, C_i and C_l takes place. Then, depending on the voltage at V_a, the comparator activates one of the two current sources. This gives a linear rise or fall in the voltage at the input and output of the comparator. When both input voltages cross, the comparator shuts down the current source and all voltages remain at the same level.

Functionally, this circuit is identical to an OTA-based integrator, only the settling behavior is different. Whereas the OTA has to drive the filter capacitors, the comparator only has to steer the current source. So, the comparator can be made fast with low power consumption. The capacitors are efficiently charged using a current source. In this implementation, the current sources draw only the current necessary to change the integrated charge. There is no waste of charge.

However, this implementation is very sensitive to comparator delay. In a CBSC system, comparator delay leads to an overshoot of the output voltage. In this implementation, this overshoot is sometimes positive and sometimes negative, depending on which current source was activated. Moreover, if both current sources are not perfectly matched, one of the two overshoots will be larger.

In the implementation shown in Fig. 2b, the integration phase is split in two. First, a short precharge takes place. During this phase, the output is shorted to a low preset voltage (V_p). As the total charge on node V_a cannot change, the integrated value is not lost. By properly choosing V_p, it can be assured that V_a will always be below the comparator threshold (V_{th}) after the preset phase. Then, the current source is activated until the comparator detects a threshold crossing of its inputs. At that moment, $V_a = V_{th}$ and all charge is transferred to C_i. The current source is shut down and all voltages are held constant.

Overshoot due to a finite comparator delay still exists, but it is always the same fixed value. Consequently, it will give a DC offset, but no distortion. When implementing this type of CBSC integrator, the voltage swing at V_a has to be carefully analyzed. Depending on the integrator gain coefficient, the integrator input and output swing and the preset voltage (V_p), V_a can fall below V_{ss} during the preset phase. It must be ensured that no pn-junction of any switch becomes forwardly biased. This would leak charge away, resulting in a loss of the integrated value.

The implementation shown in Fig. 2c, further splits up the loading phase (Φ_{2l}) [6]. First, a large current source is activated, leading to a swift change in the output voltage and a large overshoot. Then, a smaller current source is activated to correct the overshoot. Because of the small current source, a larger comparator delay can be allowed for the same overshoot. However, due to the requirements on non-overlapping clocks, this implementation is less suited for high sampling frequencies.

The CBSC integrators shown in Fig. 2 are single-ended. However, operation with differential signals is preferable. As it quadruples the signal swing and doubles the noise, all capacitors can be halved for the same SNR. Moreover, the A/D converter becomes less susceptible to power supply and substrate noise. Figure 3 shows

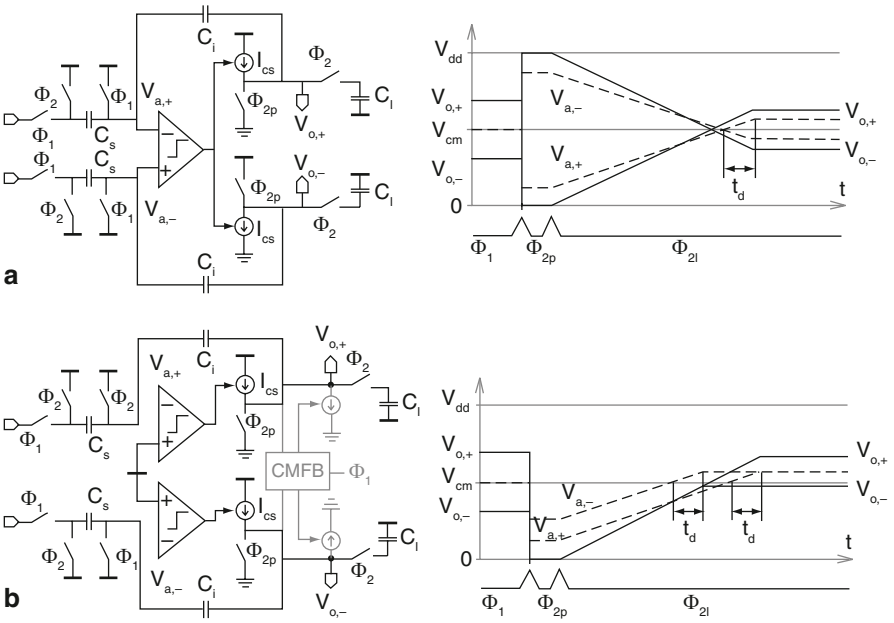


Fig. 3 Fully-differential and pseudo-differential implementation of a CBSC integrator with preset phase. **a** Fully-differential implementation. **b** Pseudo-differential implementation

two implementations of a CBSC integrator with preset that operate on differential signals.

The first implementation is fully-differential and has only one comparator. The two outputs are preset to opposite supply rails. Then, two different current sources are activated, making the two output voltages (and consequently the comparator input voltages) evolving towards each other. When the two comparator input voltages cross, the current sources are turned off. Comparator delay (t_d) causes the current sources to be on for a too long time. This results in a differential offset voltage at the output.

The second implementation, shown in Fig. 3b, is pseudo-differential. The two outputs are preset to the same voltage and identical current sources can be used to charge C_i. As a result, the comparator delay (t_d) results in a common-mode offset of the output voltage. This can easily be corrected with a common-mode feedback circuit. Such a common-mode feedback can operate during the other clock phase (as shown in Fig. 3b). This only corrects the common-mode error on C_i, but in a cascade of integrators, C_l is the sampling capacitor of the next integrator. Consequently, the common-mode error on C_l is corrected by that integrator. Since the common-mode error does not accumulate on C_l, this poses no problem. The advantage of this approach is that an entire clock phase is available for this common-mode correction. This allows the use of small current sources, leading to a small residual error.

As a common-mode error is easier to correct than a differential error, the pseudo-differential implementation puts less stringent requirements on t_d. Therefore, its power consumption is not necessarily much higher than that of a fully-differential implementation.

3 CBSC: Comparator

This section discusses the impact of a non-ideal comparator. The comparator must translate an input threshold crossing into a change of its output logic value. Therefore, it is called a threshold crossing detecting comparator (TCDC). This in contrast with a dynamic comparator as often used in the quantizer [2, 3]. Such a comparator must determine a logic value at a specified time instant. This section treats comparator delay, comparator offset and comparator noise. In Sect. 3.4, a further simplification of this comparator is explained.

3.1 Comparator Delay

Comparator delay (t_d) results in an overshoot of the output voltage. This overshoot is given by (1). C_t is the total load capacitance during the integration phase, given by (2). For a single-ended or a fully-differential implementation, this overshoot results in a DC offset of the output signal. For a pseudo-differential implementation, the delay gives a common-mode error, which is easier to correct.

$$V_{ov,o} = \frac{I_{cs}}{C_t} t_d \tag{1}$$

$$C_t = \frac{C_s \cdot C_i}{C_s + C_i} + C_l \tag{2}$$

A TCDC can be implemented by cascading several gain stages. Such a gain stage can be modeled with a finite gain A_0 and a single pole at $f_0 = s_0/(2\pi)$. Because the comparator acts in open loop, no stability issues arise when cascading multiple stages. The more stages used, the higher the total gain will be, but the delay from input to output also rises. However, the comparator does not need complete settling. The only important figure is the delay between an input threshold crossing and the moment that the comparator output flips. This delay is independent of the filter capacitors C_s and C_i. This is contrary to an OTA-based system, where settling is determined by C_s. The response of a single-pole stage on a constant slope input is given by (with $\tau = 1/s_0$):

$$V_o(t) = A_0 \frac{dV_a(t)}{dt} \left(t - \tau \left(1 - e^{-t/\tau}\right)\right) \tag{3}$$

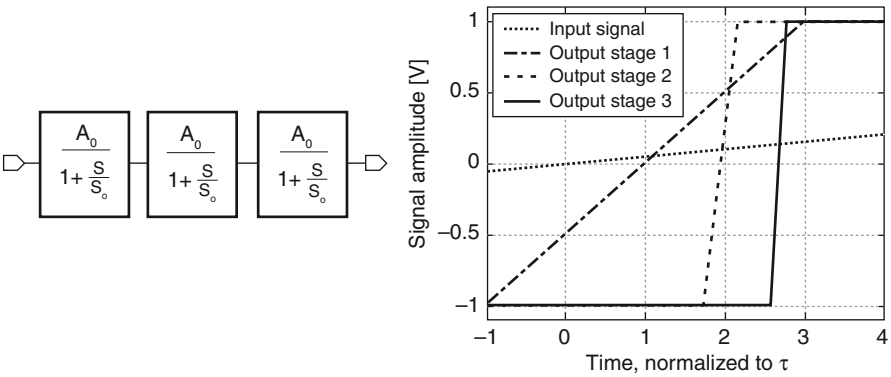


Fig. 4 A TCDC composed of three cascaded gain stages and the response to a constant slope input signal

The stage output voltage rises with the stage gain times the input slope, corrected for the effect of the pole. Consequently, if two stages are cascaded the output will rise with $(A_0)^2$ times the input slope. Because this leads to a faster changing output voltage, cascading more stages can result in a quicker transition of the comparator output voltage.

In Fig. 4, the behavior of a TCDC composed of three single-pole gain stages is simulated. Each stage has a gain of 20 dB and an input slope of 50 mV/τ is applied. The outputs are clipped between −1 V and +1 V. For this input slope, the output of the second stage is the fastest to complete the low to high transition. This transition happens in less than 2τ after the input threshold crossing. The third stage does not offer an additional benefit as the output slope of the second stage is already limited by its own time constant.

3.2 Comparator Offset

A comparator with a differential pair input has an offset due to mismatch of the two input transistors. This offset will alter the moment of the threshold detection, leading to a change in the duration that the current source is activated. The effect of an input referred offset voltage (V_{os}) on the final integrator output voltage can be calculated:

$$\Delta V_o = \frac{C_s + C_i}{C_i} V_{os} \tag{4}$$

In a pseudo-differential implementation, both comparators have an offset. This results in a common-mode effect and a differential effect on the output voltage. In a good design, transistor matching can be made fairly accurate, making the effect of

offset much smaller than the effect of delay. If nevertheless the differential error is too large, auto-zeroing or chopping techniques can be introduced [10].

3.3 Comparator Noise

The noise of the comparator will influence the moment of the threshold crossing detection. If the comparator noise is represented as an input referred voltage source, the threshold level becomes noisy, resulting in variation of the moment that the comparator detects a threshold crossing. This is equivalent to jitter on the comparator delay.

For a CBSC system that operates with a preset phase and a shut down of the current sources in less than a few time constants, non-stationary noise analysis needs to be applied [6, 7]. After the preset, the comparator current noise starts to integrate until the moment that the current source is shut down.

For a comparator composed of a cascade of gain stages, only the noise of the first stage is of importance. The input referred noise of the subsequent stages is suppressed by the first stage gain. Consequently, the other stages can be modeled as an ideal TCDC. The output voltage noise power of the first stage (with transconductance g_m and noise excess factor γ) results from the integration of its white current noise on its output capacitor (C_{out}):

$$\overline{v_{n,out}^2(t)} = \frac{4kT(2/3)g_m\gamma}{2C_{out}^2}t \tag{5}$$

To calculate the equivalent input referred noise, the noise gain of the stage has to be calculated. The noise results in jitter on the comparator threshold detection moment. The amount of jitter is determined by the output slope of the gain stage. Consequently, the noise gain is the ratio of the output slope of the first stage (given by the derivative of (3)) to its input slope. This output slope rises with time. For $t < \tau$, the noise gain becomes:

$$A_N(t) = \frac{g_m}{C_{out}}t \tag{6}$$

From this, the equivalent input referred noise power at moment of the threshold crossing (t_d) results:

$$N_{comp} = 4kT\frac{2/3}{g_m}\gamma\frac{1}{2t_d} \tag{7}$$

Because the noise gain rises with time, the input referred noise of the comparator becomes lower with a larger comparator delay. However, the overshoot increases and the current source noise becomes higher for a larger t_d, as shown in Sect. 4.3.

Consequently, for a CBSC implementation with a small current source, as shown in Fig. 2c, a large comparator delay can be advantageous [6]. When larger current sources are used, it is best to minimize the comparator delay.

3.4 Zero-Crossing Based Implementation

For all integrator implementations shown in Fig. 2, the input of the comparator is a constant slope voltage ramp. The comparator itself has to detect when this voltage crosses a certain threshold voltage. For this function, a general purpose comparator is not always required. A very simple implementation is using the V_T of a single transistor, leading to a zero-crossing based implementation [7–9].

An integrator with preset phase using a single transistor zero-crossing detector is shown in Fig. 5. Transistor M2 is biased with V_b, chosen such that it guarantees that the current source remains turned on as long as V_a is low. When V_a crosses the threshold voltage V_T of M1, M1 starts conducting, pulling down V_c, until the current source shuts down. During Φ_1, V_b can be connected to V_{dd}, so that there is no static power consumption during this clock phase. Because of this simple structure, the power consumption can be lower than when a multiple gain stage comparator is used. Due to the simple architecture, the noise excess factor γ in (7) is also smaller.

The disadvantage of this technique is that the threshold crossing voltage now depends on transistor parameters. These change with technology, temperature and process variations and result in a DC offset. Consequently, techniques to reduce this offset need to be applied [9, 10]. Also, at the threshold crossing point, the gain of the zero-crossing detector will be relatively low in nanoscale CMOS technologies. This results in a slow transition edge at V_c, possibly leading to a gradual lowering of I_{cs} instead of an ideal on/off behavior.

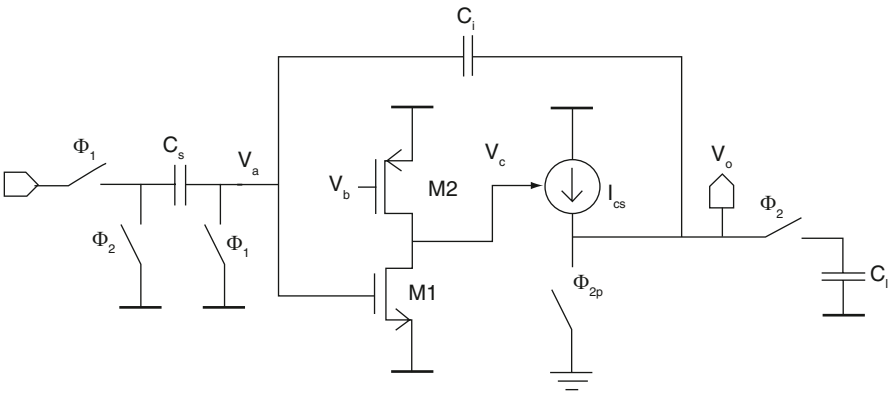


Fig. 5 A zero-crossing based switched-capacitor integrator

4 CBSC: Current Source

In this section the requirements for the current source are discussed. This comprises the sizing, the output resistance and the effect of noise. In Sect. 4.4 different implementations with a high output resistance and high output swing are discussed.

4.1 Current Source Sizing

The current delivered by the current sources is best chosen as low as possible. The lower the current, the higher the allowed comparator delay for the same overshoot voltage (1). The minimum value is determined by the charge the current source must deliver in the worst case scenario. For a CBSC integrator with a preset of the output to V_p, this results in the following condition:

$$I_{cs} \geq \frac{C_t \left(V_{o,max} - V_p\right)}{t_{\Phi_{2l}}} \tag{8}$$

4.2 Current Source Output Resistance

A real current source has a finite output impedance. This makes the delivered current dependent on the output voltage. For a current source modeled as (9), the overshoot given by (1) changes into (10).

$$I_{cs} = I_{cs,0} + \frac{V_o}{R_{cs}} \tag{9}$$

$$V_{o,ov} = \frac{I_{cs,0}}{C_t} t_d \left(1 + \frac{V_o}{I_{cs,0} R_{cs}}\right) \tag{10}$$

The first term of (10) results in a constant DC offset of the output. The second term makes the overshoot signal dependent, resulting in a non-constant error. To minimize this effect, the current source should have a large output resistance. The smaller t_d, the smaller the effect of the signal dependent overshoot.

4.3 Current Source Noise

The noise from the current source results in a random walk noise voltage on the integrator output. After the preset, the current source starts charging the output capacitors. Variations in the charging current due to noise, lead to a voltage noise which grows larger with time.

Only the part of this noise during the comparator delay (t_d) is of importance. In the beginning when the current source is turned on, its noise will change the threshold detection moment but this does not result in errors on the integrated value. Consequently, only the noise during the time between the threshold crossing and the moment the comparator shuts down the current source influences the integrated voltage. For a white current noise with power spectral density $S_{cs}(f)$, the input referred voltage noise power is given by:

$$N_{cs} = \frac{S_{cs}(0)}{C_t^2} t_d \left(\frac{C_i}{C_s + C_i} \right)^2 \tag{11}$$

The noise of the current source increases with the comparator delay. Therefore, only when a small current source, with a low $S_{cs}(f)$ is used, it is beneficial to use a slow comparator. If, for a given comparator delay, the noise of the current source is too large, it can be decreased by increasing C_t. This follows from the fact that the required current source size (8) rises linearly with C_t, whereas the noise drops with the square of C_t. However, increasing C_t increases the power consumption.

4.4 High Output Swing Implementation

A high output resistance of the current source is preferable against signal dependent overshoot. In a nanoscale CMOS technology, this means that cascoded current sources are required. However, to keep both the current source transistor and the cascode transistor in saturation, the output voltage swing seriously reduces. This can be anticipated by driving the current source from a higher than nominal supply voltage ($V_{dd,up}$). Figure 6 shows four different implementations of this principle.

The implementation shown in Fig. 6a is the most straightforward solution. The gates of the current source and the biasing transistor are kept at their biasing voltage and an additional series switch is added. This works fine when no high supply voltage is used. However, if $V_{dd,up} > V_{dd} + V_{Tp}$, the current source cannot be shut down. This can be solved by switching the gate of the switch transistor between V_{ss} and $V_{dd,up}$. However, in that case the gate-drain voltage of the switch transistor exceeds the nominal supply voltage when the switch is open, causing reliability problems. By switching the gate of the cascode transistor, the additional series switch is not needed anymore (Fig. 6b). However, the problems for turning off the current source remain. Figure 6c, d show two implementations that function correctly. Both perform the switching at the gate of the current source transistor. As long as $V_{dd,up} < 2V_{dd}$, the biasing voltages can be set in such a way that no voltage across two transistor terminals exceeds V_{dd}. This ensures reliable operation without the need for thick-oxide transistors. The advantage of the implementation shown in Fig. 6c is that the current can be accurately controlled thanks to the biasing network. However, to turn on the current source, the charge at its gate capacitor must be drained by the biasing current source. This results in slow operation or requires a large biasing current. In the implementation shown in Fig. 6d, the current source

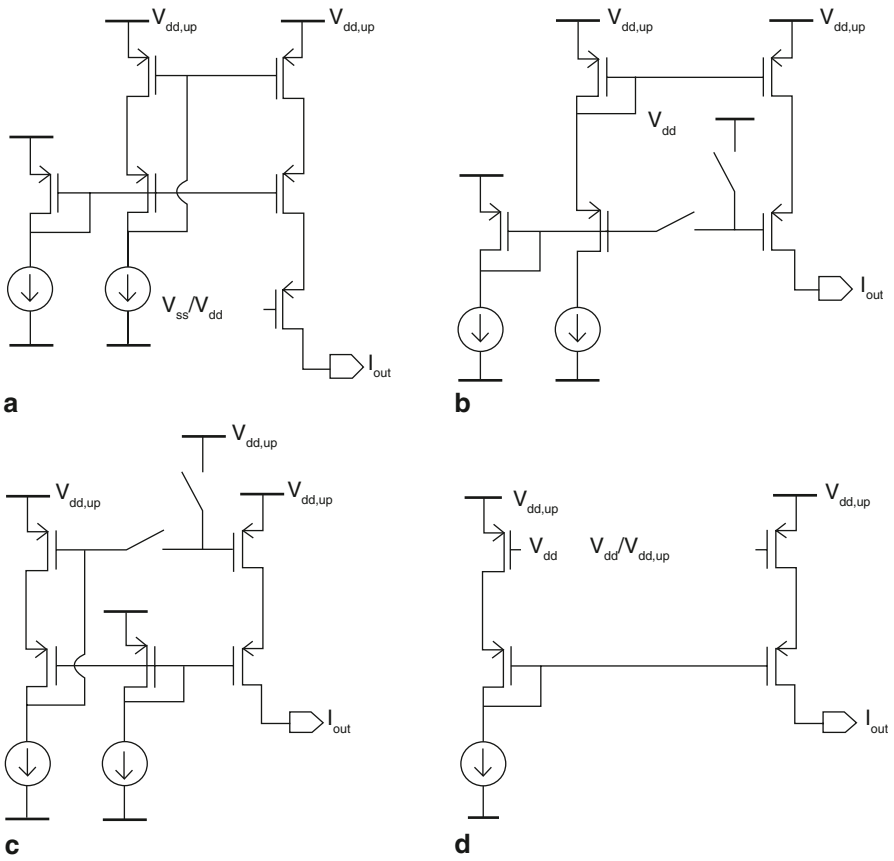


Fig. 6 Different implementations for a cascoded, high output swing current source, driven from a higher than nominal supply voltage. **a** Additional series switch. **b** Switching cascode biasing. **c** Switching current source biasing. **d** Switching current source

transistor switches between two supply voltages. This allows very fast operation. The disadvantage is that the transistor performs a voltage to current conversion and hence I_{out} is determined by transistor parameters. Small deviations in one of the two supply voltages result in large variations of this current. Therefore, sufficient bypassing capacitance is required between the two supply voltages.

5 CBSC Noise-Shaping Filter

By cascading n integrators an nth-order noise-shaping filter is obtained. All discussed CBSC integrators are half delay integrators. The output of one integrator must be sampled on the sampling capacitor of the next integrator during the

integration phase of the former. This is because the CBSC topology only makes a virtual ground when the current source has been activated and does not try to maintain it. Therefore, the circuit configuration may not be altered after the current sources are shut down without reactivating them again. Using half delay integrators does not necessarily degrade the noise-shaping transfer function, so this is no problem [11].

Feed-forward noise-shaping filters are currently very popular. They are less sensitive to non-linearities of the filter components, because the filter loop processes only quantization noise [12]. This technique requires an analog adder before the quantizer. For single-bit ΔΣ A/D converters, this adder can be implemented passively using switched-capacitors [13]. However, such an adder creates a feed-through path from one integrator output to another integrator output. For an OTA-based integrator implementation, this is no problem, as each OTA continuously forces the virtual ground condition at its input. However, in a CBSC implementation this is not true anymore. Each integrator will turn off its current source at a different time. The integrators still operating after the first one has reached the virtual ground condition will influence this ones output and virtual ground node. As a result, the integrated value on this integrator will become corrupted. Moreover, when using half delay integrators, the implementation of this adder requires the addition of analog half delay elements. This requires additional comparators and current sources, increasing the power consumption.

For these reasons, feed-back noise-shaping filters composed of half delay integrators are preferable for a CBSC implementation.

The sizing of the filter capacitors is determined by the required thermal noise floor. In a ΔΣ A/D converter, only the noise inside the signal band is of importance. The remainder is filtered out by the digital decimation filter after quantization. Consequently, when the total noise power is calculated, only a fraction (1/OSR) of it remains. As the input referred noise of all but the first integrator undergoes noise-shaping, the noise of the first integrator is the most important [1]. Besides the comparator and the current source, the series resistance of the switches forms an additional noise source. During the integration phase, the contribution is similar as that of the comparator. When designing, the switch resistance can be made small, so that its noise contribution can be neglected compared to the comparator noise. During the sampling phase, the switch results in noise. As, for good settling, the RC time constant is much smaller than the clock period, stationary noise analysis can be used. The resulting noise power is given by:

$$N_{R,eq} = \frac{kT}{C_s} \tag{12}$$

If the noise from the reference voltages and the preset voltage is neglected, the SNR due to thermal noise can be calculated. The total input referred noise power is found by adding (7), (11) and (12). Dividing by the OSR gives the total in-band noise power. The maximum signal power is determined by the reference voltage (V_{ref})

and the overload level (OL) of the noise-shaping filter topology. For a (pseudo-) differential implementation, this results in:

$$SNR_{th,diff} = \frac{\left(2 \cdot OL \cdot V_{ref}\right)^2}{2} \frac{OSR}{2\left(N_{comp} + N_{cs} + N_{R,eq}\right)} \tag{13}$$

Consequently, a possible design strategy is as follows. The noise from the switches (12) can be set low enough by properly choosing the filter capacitors. The comparator noise (7) is determined by its g_m and delay (t_d). When both are fixed, the noise from the current source can be calculated (11). If this noise power is too high, the comparator delay should be reduced or the filter capacitors need to be increased. The first solution requires more current for the comparator (its g_m needs to be increased in order not to increase its noise contribution). The second solution increases the current consumed by the current sources.

6 A CBSC Delta-Sigma A/D Converter in 90 nm CMOS

In this section the design and measurements of a CBSC ΔΣ A/D converter is described. It uses integrators with a preset phase and a single current source. The integrators are implemented pseudo-differentially as this allows operating with differential signals while minimizing the requirements for the comparator delay.

A fourth-order noise-shaping filter with a single-bit quantizer is implemented. A feed-back topology with four half delay integrators is used, as shown in Fig. 7. The filter coefficients are chosen for stability and a suitable output voltage swing for each integrator.

The transistor level implementation of the used integrators is shown in Fig. 8. The shown implementation integrates during Φ_2. For each consecutive integrator, the clock phases have to be swapped. The two TCDC1 comparators form the pseudo-differential implementation. A current source as shown in Fig. 6d is used. This combines a large output swing with a high output resistance. By switching the gate of the current source transistor, a fast turn-on and turn-off behavior is obtained. As this gate signal needs to switch between V_{dd} and $V_{dd,up}$, a level-shifter is necessary. This is realized by precharging a capacitor during the sampling phase and connecting it in series with the comparator output during Φ_2.

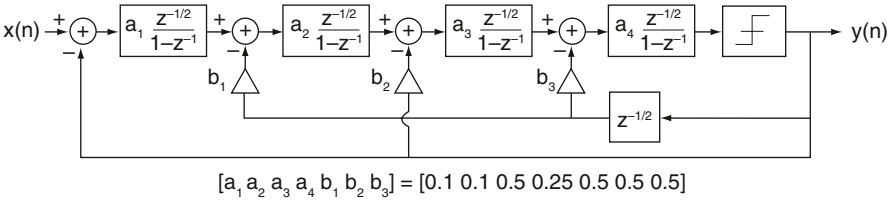


Fig. 7 A fourth-order feed-back noise-shaping filter with half delay integrators

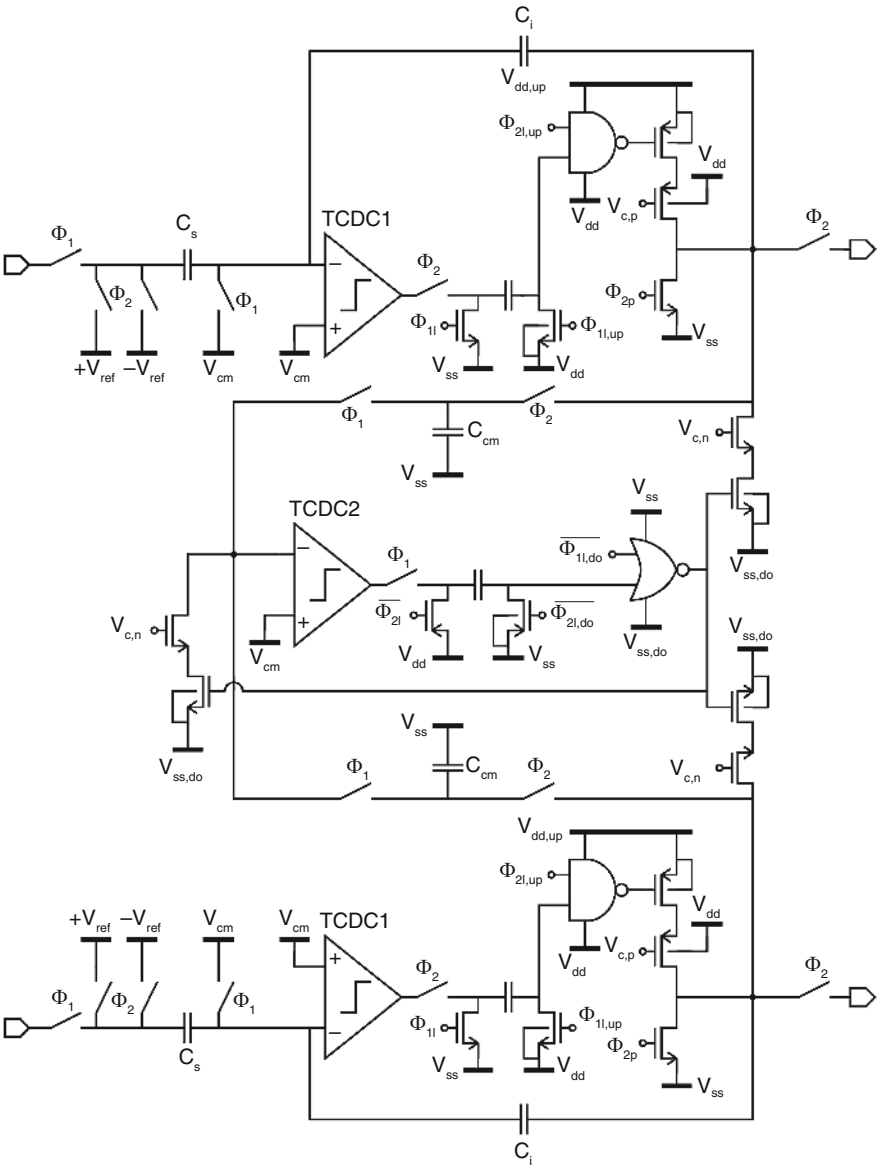


Fig. 8 Circuit level implementation of the CBSC integrator

The third comparator, TCDC2, is used for the common-mode correction. During the integration phase, both output voltages are sampled on the C_{cm} capacitors. During the next clock phase, both C_{cm} capacitors are shorted together, leading to a value for the common-mode voltage. As the integration phase used current sources to charge the output, the sampled common-mode voltage will always be

too high due to the finite delay of the TCDC1 comparators. Discharging current sources are then activated until TCDC2 detects a threshold crossing. At that moment, the common-mode voltage is at the desired level. Because an entire clock phase is available for this common-mode correction, small current sources can be used, leading to a small remaining error. The accuracy of the common-mode correction is determined by the matching between the capacitors C_i and C_{cm} and between the different current sources. These current sources are cascoded and connected to a low supply voltage ($V_{ss,do}$) to improve their linearity. This common-mode feedback circuit corrects only the common-mode error across C_i. The common-mode error on the sampling capacitor of the next integrator is not corrected. However, this is done during the common-mode correction phase of that integrator.

The transistor level implementation of the different comparators is shown in Fig. 9. All comparators in one integrator share the same biasing. Each comparator is composed of two gain stages followed by two inverters to restore the output levels. The first stage of the TCDC1 comparators is five times larger than the others, in order to have a low input referred noise power. For the TCDC2 comparator, this is not required, as its delay and noise only result in a common-mode error. The biasing current of the TCDC1 gain stages can be switched off, in order to save power during the sampling phase of the integrator. The node voltages can reset to their normal biasing setting during the integrator preset phase. Each stage has a gain of about

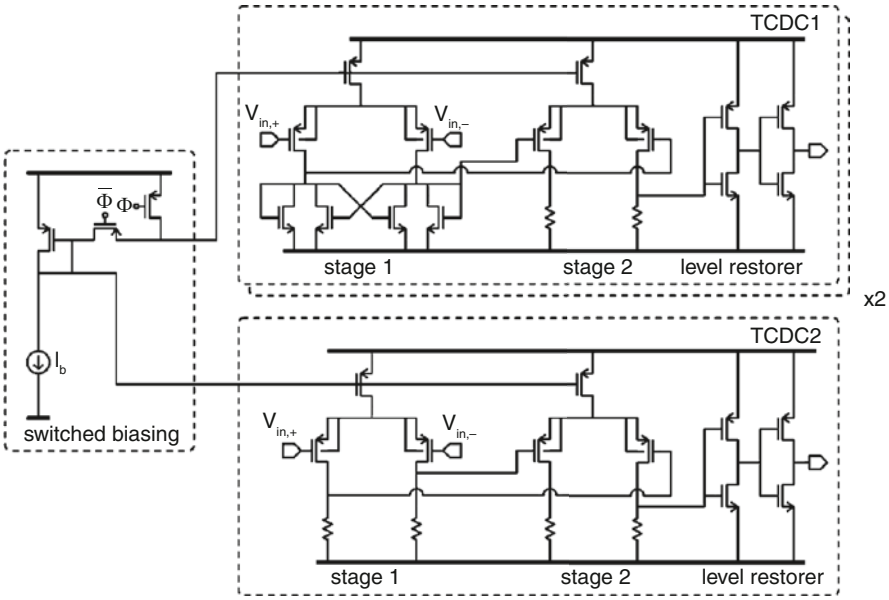


Fig. 9 Transistor level implementation of the TCDC comparators

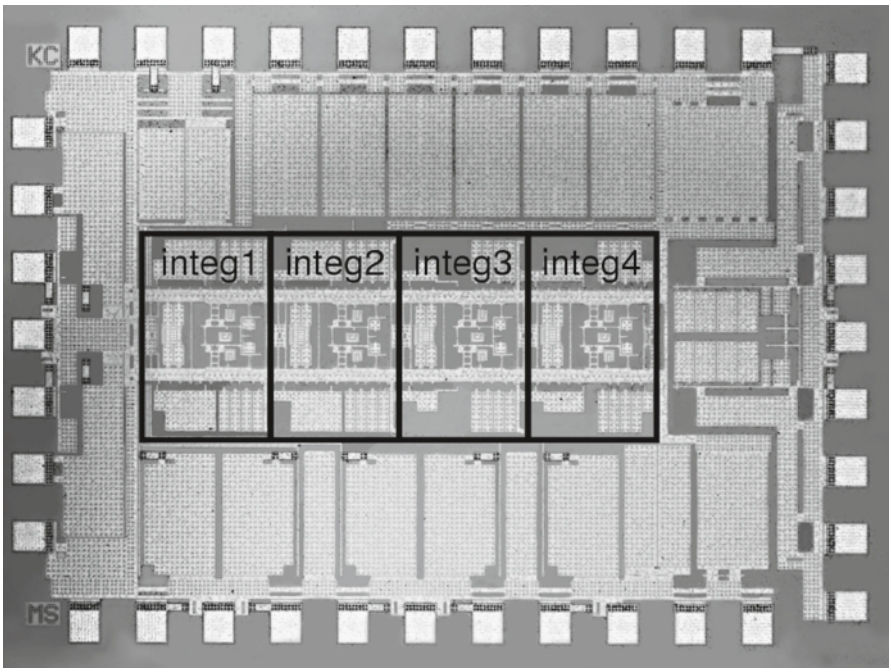


Fig. 10 Photograph of the implemented CBSC ΔΣ A/D converter

17 dB, while the first stage bandwidth is 630 MHz. The simulated comparator delay is 330 ps, which is less than 1.5τ.

A photograph of the ΔΣ A/D converter implementation in a 1 V, 90 nm CMOS technology is shown in Fig. 10. The four integrators, quantizer and clock generation circuit occupy 0.33 mm^2. The entire chip, including bonding pads and bypassing capacitors measures 1.5 mm by 1.1 mm. For measurements, this chip is bonded on a PCB. An external differential input signal and differential clock signal are applied. The output of the single-bit quantizer is captured with a logic analyzer and processed on a workstation.

The ΔΣ A/D converter is clocked at 96 MHz. For an OSR of 48, this results in a signal bandwidth of 1 MHz. Figure 11 shows the measured output spectrum for a 75 kHz sine input signal with an amplitude of 800 $mV_{ptp,diff}$. Figure 12 shows the measured SNR and SNDR when the input amplitude is swept. The dynamic range is 70 dB, while the peak SNDR is 66 dB. The ΔΣ A/D converter consumes 5.05 mW from the 1 V supply voltage. The noise-shaping filter capacitor charging current sources draw 0.55 mA from a 1.55 V supply voltage, while the common-mode correction current sources sink 0.12 mA in a −0.36 V supply voltage. This results in a total power consumption of 5.95 mW.

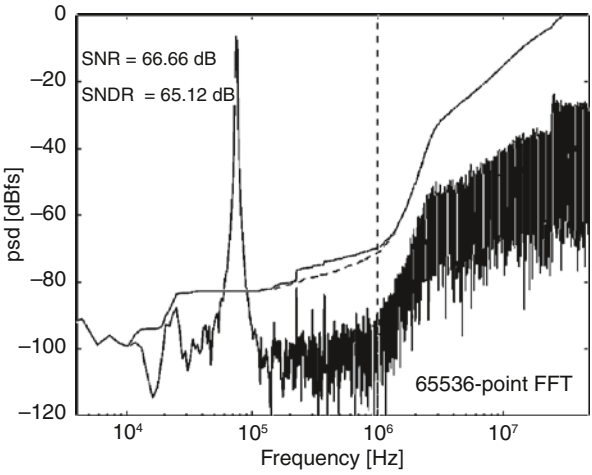


Fig. 11 Measured output spectrum for a 75 kHz, 800 mV_{ptp} sine input signal

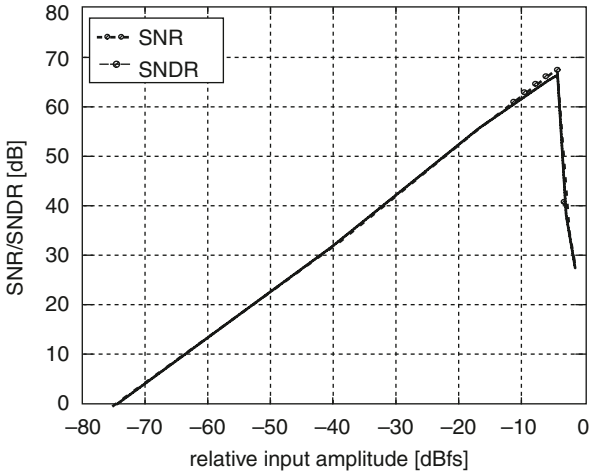


Fig. 12 Measured SNR and SNDR as function of input signal amplitude

These measurements are done without switching of the comparator biasing. When activating the comparator biasing switching, the SNDR drops to 63 dB. Because the current I_b in Fig. 9 was chosen small, the turn on behavior of the comparators becomes too slow. Therefore, for these measurements the biasing current was increased. As a result, the power consumption from the 1 V supply voltage when comparator biasing switching is activated drops only to 3.8 mW.

The performance of this CBSC ΔΣ A/D converter is summarized in Table 1. With a FOM of 1.82 pJ/conv the measured performance is comparable with other state-of-the-art 1 V ΔΣ A/D converters [2–4, 13].

Table 1 Performance summary of the implemented CBSC ΔΣ A/D converter

	Without comparator biasing switching	With comparator biasing switching
Signal bandwidth	1 MHz	1 MHz
OSR	48	48
Sampling frequency	96 MHz	96 MHz
DR	70 dB	68 dB
SNR	68 dB	67 dB
SNDR	66 dB	63 dB
Technology	90 nm standard CMOS	90 nm standard CMOS
Core area	0.33 mm^2	0.33 mm^2
V_{dd}	1 V	1 V
$V_{dd,up}$	1.55 V	1.55 V
$V_{ss,do}$	−0.36 V	−0.36 V
Input range	1 $V_{ptp,diff}$	1 $V_{ptp,diff}$
Overload level	0.75	0.75
Power consumption V_{dd}	5.05 mW	3.8 mW
Power consumption $V_{dd,up}$	0.85 mW	0.85 mW
Power consumption $V_{ss,do}$	0.04 mW	0.04 mW
Total power consumption	5.94 mW	4.69 mW
Figure of Merit	1.82 pJ/conv	2.03 pJ/conv

7 Conclusions

Comparator-based switched-capacitor circuits replace OTAs by comparators and current sources. This is advantageous for nanoscale CMOS technologies, which pose severe problems for high gain OTA design.

Different implementations of a CBSC integrator are discussed. A pseudo-differential implementation with a preset phase and a single current source is most interesting for applications with high sampling frequencies. The pseudo-differential topology translates the effect of comparator delay into a common-mode error. This error can be corrected during the next clock phase. To integrate a current source with a high output resistance and a large integrator output swing, the use of a cascoded topology operating at a higher than nominal supply voltage is proposed.

To construct a noise-shaping filter, feed-back topologies are preferable. These are easily implementable with half delay integrators and do not suffer from possible parasitic feed-through paths from one integrator to another.

To demonstrate the feasibility of a CBSC ΔΣ A/D converter, an implementation in a 1 V, 90 nm CMOS technology is made. It implements a fourth-order noise-shaping filter and operates at a sampling frequency of 96 MHz. For a signal bandwidth of 1 MHz, a peak SNDR of 66 dB is obtained. The total power consumption is 5.94 mW.

References

1. Y. Geerts et al., A 3.3-V 15-bit, delta-sigma ADC with a signal bandwidth of 1.1 MHz for ADSL applications. IEEE J. Solid-State Circuits. **34**(7), 927–936 (1999)
2. L. Yao, M. Steyaert, A 1-V 140-μW 88-dB audio sigma-delta modulator in 90-nm CMOS. IEEE J. Solid-State Circuits. **39**(11), 1809–1818 (2004)
3. K. Cornelissens, M. Steyaert, A 1-V 84-dB DR 1-MHz bandwidth cascade 3-1 Delta-Sigma ADC in 65-nm CMOS, in *Proceedings of ESSCIRC*, 2009, pp. 332–335
4. K.-P. Pun, S. Chatterjee, P. Kinget, A 0.5 V 74 dB SNDR 25 kHz CT ΔΣ modulator with return-to-open DAC, in *IEEE International Solid-State Circuits Conference Digest of Technical Papers*, Feb 2006, pp. 181, 182
5. Y. Chae, G. Han, Low voltage, low power, inverter-based switched-capacitor Delta-Sigma modulator. IEEE J. Solid-State Circuits. **44**(2), 458–472, Feb (2009)
6. J.K. Fiorenza et al., Comparator-based switched-capacitor circuits for scaled CMOS technologies. IEEE J. Solid-State Circuits. **41**(12), 2658–2668 (2006)
7. L. Brooks, H.S. Lee, A zero-crossing-based 8-bit 200 MS/s pipelined ADC. IEEE J. Solid-State Circuits. **42**(12), 2677–2687 (2007)
8. S.K. Shin et al., A fully-differential zero-crossing-based 1.2 V 10 b 26 MS/s pipelined ADC in 65 nm CMOS, in *IEEE Symposium on VLSI Circuits*, June 2008, pp. 218, 219
9. L. Brooks, H.S. Lee, A 12 b 50 MS/s fully differential zero-crossing-based ADC without CMFB, in *IEEE International Solid-State Circuits Conference Digest of Technical Papers*, Feb 2009, pp. 167, 168
10. C.G. Enz, G.C. Temes, Circuit techniques for reducing the effects of op-amp imperfections: autozeroing, correlated double sampling, and chopper stabilization. Proc. IEEE. **84**(11), 1584–1614 (1996)
11. V. Peluso et al., A 900 mV 40 μW switched opamp ΔΣ modulator with 77 dB dynamic range, in *IEEE International Solid-State Circuits Conference Digest of Technical Papers*, Feb 1998, pp. 68, 69
12. J. Silva, U. Moon, J. Steensgaard, G.C. Temes, Wideband low-distortion delta-sigma ADC topology. Electron. Lett. **37**(12), 737, 738 (2001)
13. L. Yao, M. Steyaert, W. Sansen, A 1-V 1-MS/s, 88-dB, sigma-delta modulator in 0.13-μm digital CMOS technology, in *Symposium on VLSI Circuits Digest of Technical Papers*, June 2005, pp. 180–183

VCO-Based Wideband Continuous-Time Sigma-Delta Analog-to-Digital Converters

Michael H. Perrott

1 Introduction

Continuous-time Sigma-Delta ADC structures have recently emerged as a promising candidate to achieve wide bandwidth and high SNDR analog-to-digital conversion with reduced anti-alias filtering requirements compared to their discrete-time ADC counterparts [1–15, 16]. As shown in Fig. 1, a continuous-time Sigma-Delta ADC avoids front end sampling and instead utilizes a feedback loop that seeks to track the low frequency content of its input with the output of a DAC circuit. By clocking the DAC output at a high rate (i.e., oversampling the bandwidth of interest), the effective tracking resolution can be made quite high at low frequencies so that a high SNDR is achieved for the ADC within the bandwidth of interest.

While the DAC structure is most simply implemented as a 1-bit structure, this choice will lead to very high oversampling ratios when seeking high SNDR and may prove to be impractical when high signal bandwidth is desired. With the advent of dynamic element matching techniques [17], there has been a growing trend to use multi-bit DAC circuits within continuous-time Sigma-Delta ADCs in order to lower such oversampling requirements and simplify stability analysis.

Assuming a multi-bit DAC is utilized, the quantizer must also have a multi-bit implementation as indicated in Fig. 1. To achieve a wideband ADC, the quantizer and DAC must be clocked at a high sample rate in order to provide an adequately high oversampling ratio. In order to achieve stable feedback dynamics, delay in the quantizer operation should be minimized.

In this chapter, we will focus on achieving efficient implementation of a multi-bit quantizer using a voltage-controlled ring oscillator (VCO) [18, 19, 20–25]. We will show that the resulting structure can be viewed as an efficient combination of a Voltage-to-Time Converter (VTC) and Time-to-Digital Converter (TDC), and that it can achieve high sample rates with acceptable delay and relatively low area and power. However, a key drawback of the VCO-based quantizer is nonlinearity,

M. H. Perrott (✉)
SiTime Corporation, Sunnyvale, USA
e-mail: mhperrott@gmail.com

H. Casier et al. (eds.), *Analog Circuit Design,*
DOI 10.1007/978-94-007-0391-9_10, © Springer Science+Business Media B.V. 2011

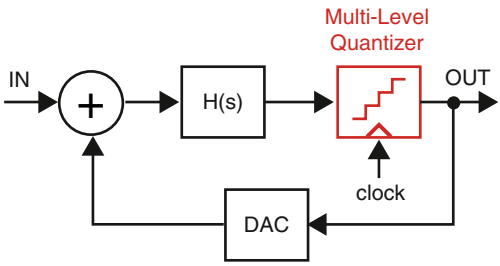


Fig. 1 A continuous-time Sigma-Delta ADC utilizing a multi-level quantizer

which can hamper the achievement of high SNDR. We will examine the impact of this nonlinearity, and then show that its influence can be largely removed by making use of phase rather than frequency as the key variable in the VCO-based quantizer output. Measured results of several 20 MHz bandwidth, continuous-time Sigma-Delta ADC prototypes using VCO-based quantization are provided, where 67 dB SNDR is demonstrated for a simple 3rd order ADC structure using frequency as its key variable, and 78 dB SNDR for a more sophisticated 4th order ADC structure using phase as its key variable.

2 Background on VCO-Based Quantization

Time-to-Digital Converters (TDC) have recently become a topic of interest in achieving ADC implementations with higher digital content, thus enabling them to better leverage Moore's law to achieve low power and area [26, 27]. As a perspective on this issue, we offer the viewpoint that VCO-based quantization can be considered as an efficient approach of achieving combined voltage-to-time and time-to-digital conversion. To explain this viewpoint, we will first review principles of time-to-digital conversion. We will then examine the VCO as a means of performing voltage-to-time conversion (VTC). By combining the VTC with a TDC, a quantizer is created with voltage as its input and a digital code as its output. A simple model of the VCO-based quantizer is then presented which reveals its ability to perform noise shaping, and also indicates the performance impact of VCO phase noise and Kv nonlinearity. Finally, we summarize useful properties of the VCO-based quantizer that make it an interesting candidate for use within continuous-time Sigma-Delta ADC structures.

Figure 2 displays a basic TDC that measures the time difference between edges of the input signals *tin(t)* and *clk(t)* in increments of time set by the delay per stage of a buffer chain. This TDC is a digital circuit that essentially performs the analog operation of quantizing a signal, but uses time as the signal domain rather than voltage or current. A digital quantizer has the attractive aspect of directly benefiting from Moore's law since area, power, and buffer delay (which corresponds to quantizer resolution) are steadily reducing with new CMOS processes. Due to these benefits, TDCs have recently become a subject of active research [28–33].

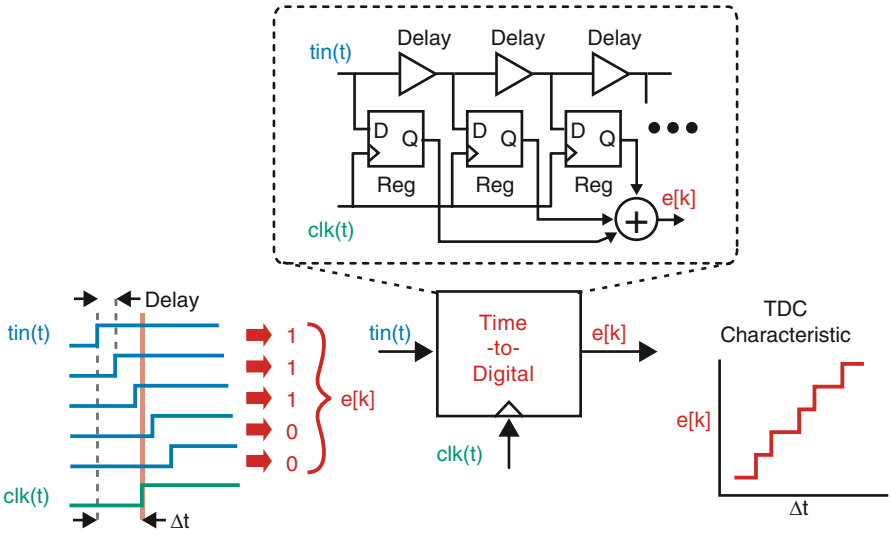


Fig. 2 Time-to-digital converter based on a single delay chain

Since an ADC must ultimately work with voltage signals, a TDC will not prove useful for this application unless there is some means of converting the voltage signal to a time-based signal. While many voltage-to-time converters are possible, one simple approach is shown in Fig. 3. Here a VCO is used to translate the input voltage into a series of edges whose period depends on that voltage. A TDC can then be used to measure the time difference between edges, which yields a digital output code that is proportional to the input voltage of the VCO. Therefore, the VCO and TDC combination provides a simple means of achieving an ADC. One should note that other methods also exist to perform voltage-to-time conversion [26, 27]. In

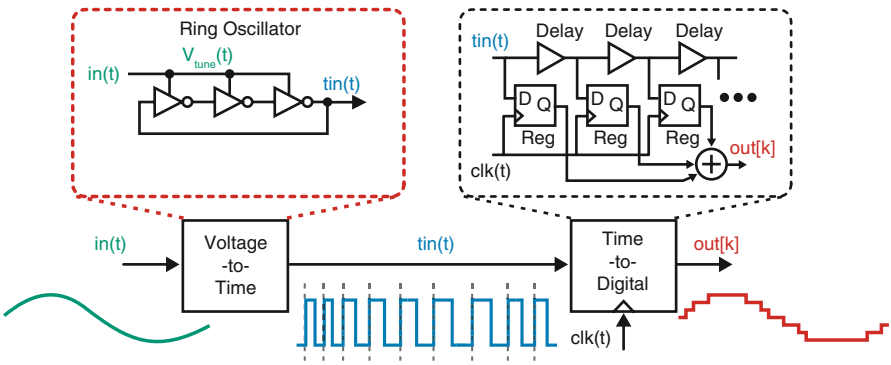


Fig. 3 A voltage quantizer based on a TDC which uses a VCO to perform voltage-to-time conversion

comparison to these other techniques, the quantizer shown in Fig. 3 has the advantage of providing a simple, highly digital implementation.

Unfortunately, the simplicity of the quantizer in Fig. 3 is also accompanied by a few shortcomings. Aside from issues of noise and nonlinearity, as we will discuss shortly, the TDC operates on edges in a manner which leads to non-uniform sampling of the input. This non-uniform sampling leads to distortion of the input signal. One way around this issue is to include a sampler in front of the overall quantizer [26], but this approach introduces extra analog complexity. A different approach is to use algorithms to deal with the non-uniform sampling [27], but this introduces significant time delay in the quantizer which would not be suitable within a feedback system.

Figure 4 illustrates an interesting variation of the TDC quantizer implementation which yields a simpler design and reduces the impact of non-uniform sampling. Rather than constraining the ring oscillator output to only one of its stages, all of the stages of the ring oscillator are utilized by feeding their outputs directly into the registers of the TDC. By doing so, the TDC samples the phase state of the ring oscillator at each clk edge rather than the time error between the output of one oscillator stage and a corresponding clk edge.

As shown in Fig. 4, the utilization of all of the ring oscillator stages leads to a much more uniform sampling of the input. To elaborate on this point, note that each oscillator delay stage essentially samples the input as an edge travels through it, and the time location of its edges are a function of the accumulative impact of input samples across all stages. Therefore, one can consider the time between edges of a given oscillator stage as a time window over which input samples are averaged. When only one oscillator edge is measured per TDC clk period, this time window shifts dramatically relative to the TDC clk edge. However, when all of the oscillator edges are utilized, the time window is much more aligned to the TDC clk edge, yielding more uniform sampling of the input.

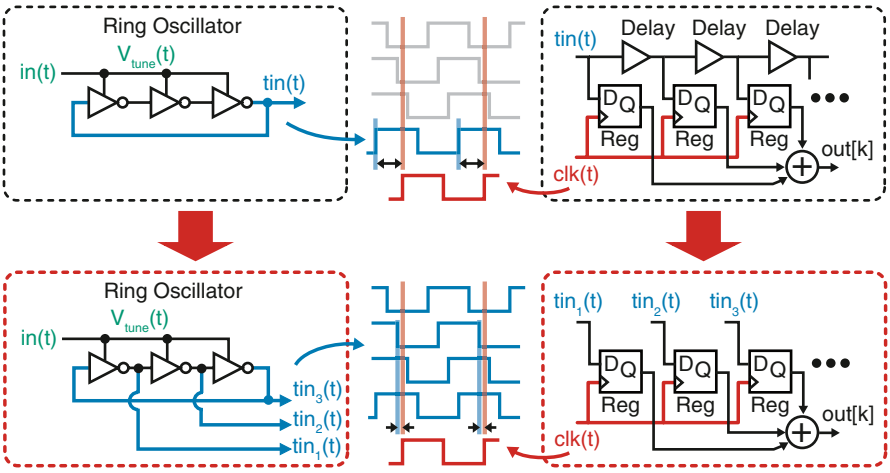


Fig. 4 VCO-based quantizer seen as a combined ring oscillator and TDC

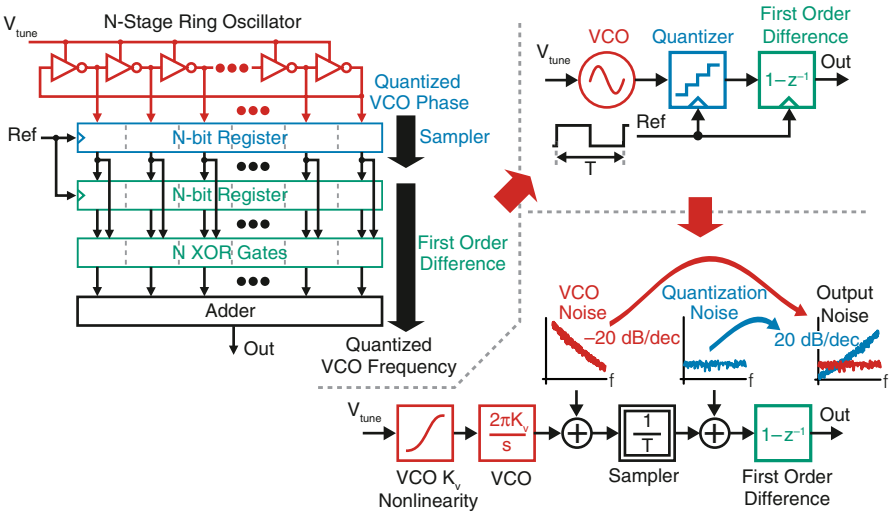


Fig. 5 High speed VCO-based quantizer with accompanying models

Figure 5 displays an efficient high-speed implementation of the quantizer illustrated in Fig. 4 combined with additional digital processing logic [14, 18–19, 34]. Here the N stages of a ring oscillator, whose frequency is controlled by the input voltage $V_{tune}(t)$, are fed into a set of sampling registers that are clocked by the *Ref(t)* signal. The outputs of the sampling register are compared to their previous values using a set of XOR gates, and the number of resulting "1" values are added to form the overall output.

A simplified block diagram of the high-speed quantizer is shown in the top, right portion of Fig. 5. As shown by this diagram, the sampling registers form a quantizer, and the XOR gates perform a first order difference operation based on the constraint that the VCO phase is always increasing [14].

Finally, a frequency domain model of the high-speed quantizer is shown in the bottom, right portion of Fig. 5. This model illustrates that the VCO acts to translate the input voltage to a phase signal which is sampled and quantized by the register. The first order difference operation of the XOR gate converts the quantized VCO phase to a corresponding frequency signal. A key benefit of the first order difference operation is that it provides noise shaping of both the VCO phase noise and the quantizer error introduced by the sampling register. As such, VCO phase noise ends up dominating at low frequencies and quantization noise ends up dominating at high frequencies. Unfortunately, the model also points out that a practical VCO-based quantizer suffers from nonlinearity in its voltage-to-frequency characteristic. As we will see, the impact of such nonlinearity presents an obstacle when seeking an ADC with high SNDR.

Figure 6 reveals additional details of the high-speed VCO-based quantizer's operation over several *Ref* clock periods. The quantizer output corresponds to the

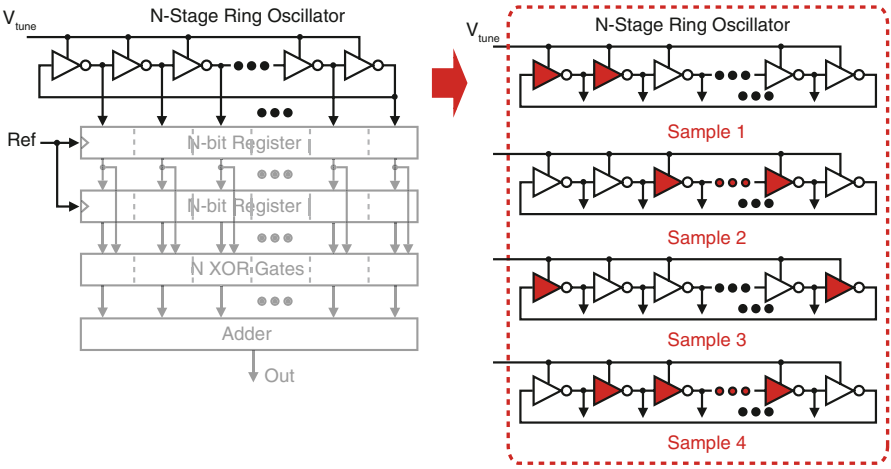


Fig. 6 Barrel shifting property of VCO-based quantizer

number of delay stages that an edge passes through within a given *Ref* clock period, as indicated by the shaded stages shown in Fig. 6. When V_{tune} increases the VCO frequency, edges travel through more delay stages in a given *Ref* clock period so that the instantaneous quantizer output increases in value. Vice versa, when V_{tune} decreases the VCO frequency, edges travel through fewer delay stages in a given *Ref* clock period so that the instantaneous quantizer output decreases in value.

Note that a key constraint on the high speed VCO-based quantizer structure is that the *Ref* clock must have sufficiently high frequency to avoid having VCO edges travel beyond *N* delay stages within a given *Ref* clock period. If this condition is violated, the edge count measured during that *Ref* clock period will correspond to the true edge count modulus *N*. While additional wrapping circuits can be added to correct for this issue, the added complexity, area, and power required for such circuits is undesirable.

A final point to observe from Fig. 6 is that the delay stages of the high speed VCO-based quantizer are visited by the edges in barrel shift fashion. As is well known in the ADC community, barrel shifting through elements has the effect of providing first order shaping of their mismatch [17]. The benefit of this property will be discussed later in this paper when we turn our attention to placement of the quantizer within the overall ADC structure.

We conclude our background discussion of VCO-based quantization by providing a simple comparison to more traditional analog flash quantizers. Figure 7 illustrates a classical analog flash converter which uses a resistor ladder to create incremental voltage comparison values and pre-amplifiers and comparators to perform the quantization operation. The pre-amps reduce the impact of comparator offset, provide reverse isolation for clock feedthrough on the comparators, and improve the metastability behavior of the quantizer. However, the resistor ladder and pre-amps demand power and area, especially if high speed operation is desired.

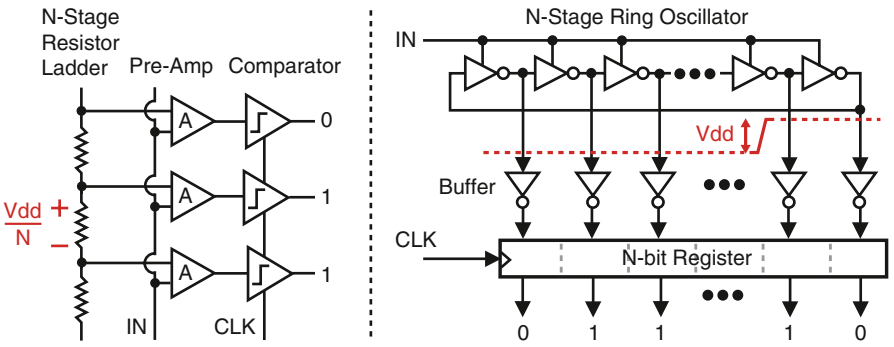


Fig. 7 Comparison of VCO-based quantizer to a classic analog flash quantizer

In contrast, a VCO-based quantizer is mostly digital in its implementation. Rather than using a resistor ladder, incremental comparison values are formed in the time domain by the delay stages of the oscillator. Since the comparison signals consist of edges rather than voltage waveforms, simple inverters are utilized instead of pre-amps. Metastability behavior is not as significant of a concern since all ring oscillator stages will have outputs at supply or ground except for the stage in transition. Finally, as discussed earlier, the VCO-based quantizer offers noise shaping of its quantization error by virtue of the first order difference operation performed by the XOR gates. Also, the barrel shifting action that occurs through its elements leads to first order shaping of the mismatch between its stages. One should note that the classical analog quantizer lacks each of these noise shaping properties.

3 A First Pass at Utilizing VCO-Based Quantization Within Continuous-Time Sigma-Delta ADC Structures

Now that we have provided background on VCO-based quantization, we turn our attention to utilizing this circuit within a continuous-time Sigma-Delta ADC. At first glance, this is quite straightforward—we simply place the quantizer within the feedback loop of the ADC. However, as shown in Fig. 8, we can go one step further and connect the unit elements of the feedback DAC directly to the individual register bits of the VCO-based quantizer. By doing so, we avoid the need to explicitly add the quantizer bits, thus saving digital logic in the quantizer implementation. Further, in this arrangement, the barrel shifting property of the VCO-based quantizer is transferred to the unit elements of the DAC. This barrel shifting action leads to Dynamic Element Matching (DEM) of the DAC elements, which greatly reduces their matching requirements. As illustrated in Fig. 8, the DEM operation formed by the VCO quantizer is achieved implicitly with no extra hardware, no additional latency, and no extra power or area [35].

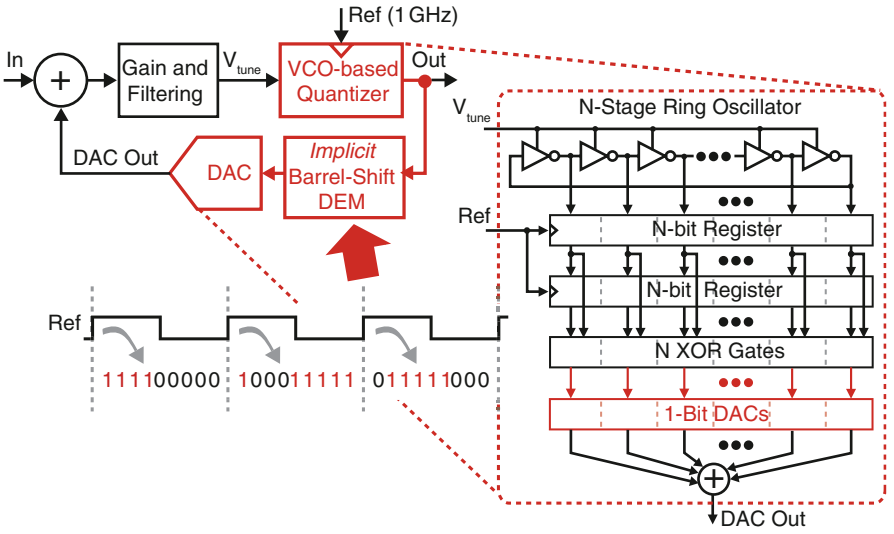


Fig. 8 A continuous-time Sigma-Delta ADC using a VCO-based quantizer

In addition to its ability to provide an implicit DEM operation, we can also leverage the quantization noise shaping property of the VCO quantizer to further simplify the ADC implementation. To illustrate this point, Fig. 9 shows a simple continuous-time ADC implementation that consists of a VCO-based quantizer, two feedback current DACs (I_{DAC1} and I_{DAC2}), and a single opamp-based feedback circuit [14]. The open loop transfer function of this structure consists of one integrator and zero provided by the opamp feedback circuit, and a second lossy integrator formed by feeding the current of I_{DAC1} into capacitance at node V_A. As such, this structure has second order dynamics, which would normally lead to second order noise shaping.

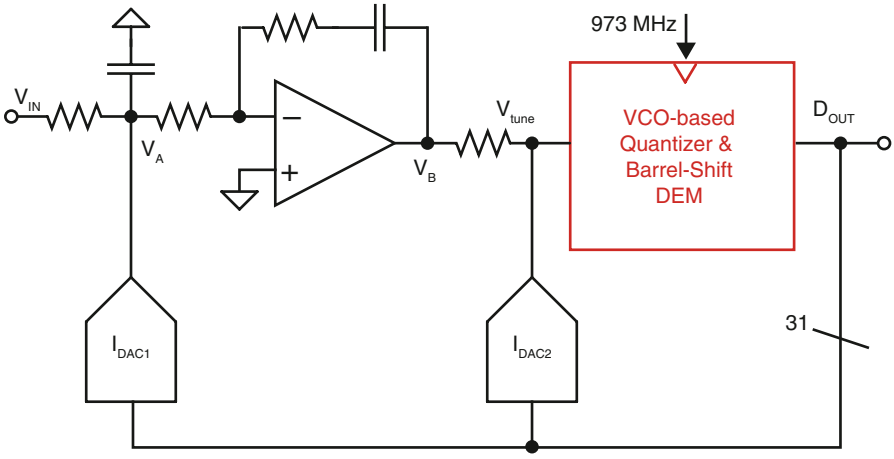


Fig. 9 A simple continuous-time Sigma-Delta ADC topology offering third order quantization noise shaping with only one opamp

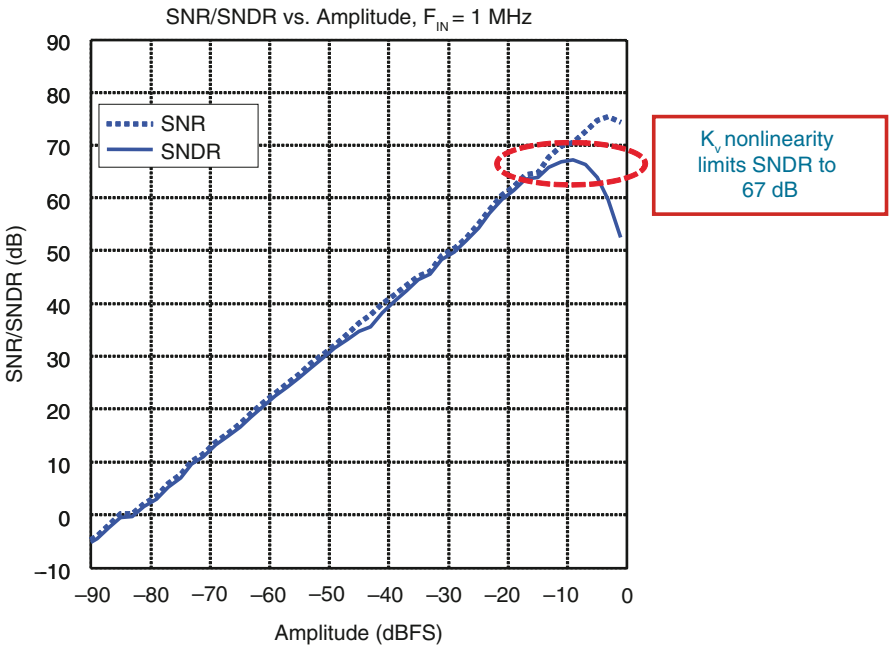


Fig. 10 Measured SNR and SNDR performance of the ADC in Fig. 9 assuming 20 MHz signal bandwidth with 973 MHz sample rate

However, we in fact achieve third order noise shaping from this ADC topology due to the additional noise shaping provided by the VCO-based quantizer, thus showing the merit of VCO-based quantization for improved quantization noise performance.

To demonstrate the practicality of the ADC structure shown in Fig. 9, a prototype was implemented in 0.13 μ CMOS. The simplicity of this ADC allows a high sample rate of 973 MHz to be achieved with relatively low power consumption of 40 mW and active area of 700 μm by 700 μm [14].

Figure 10 shows measured SNR and SNDR performance for the prototype assuming 20 MHz signal bandwidth and a 1 MHz input signal. Peak SNR approaches an impressive value of 75 dB for such a simple structure. Unfortunately, peak SNDR only reaches about 67 dB due to harmonic distortion caused by the Kv nonlinearity of the VCO-based quantizer. This example reveals that we must overcome this nonlinearity issue to improve SNDR.

4 Overcoming the Issue of Nonlinearity in the VCO-Based Quantizer

The previous section outlined some of the key benefits of using a VCO-based quantizer within a continuous-time Sigma-Delta ADC, including an increase in noise shaping order and a simple means of providing DEM of the unit elements in the

DACs. However, it also pointed out a critical bottleneck to achieving high SNDR in the overall ADC—the nonlinearity of the voltage-to-frequency characteristic of the VCO-based quantizer. While one might consider increasing analog or digital complexity to try to improve the linearity of the VCO voltage-to-frequency characteristic, it is attractive to seek a solution that retains simplicity in the overall ADC design process. While digital calibration looks promising as a means to deal with such nonlinearity [36], long calibration times may be undesirable in many applications.

Figure 11 displays an architectural approach that provides substantial reduction of the impact of nonlinearity in the VCO-based quantizer by utilizing phase information of the quantizer [15]. We begin by examining Fig. 11a, which displays the use of frequency as the key quantizer variable as assumed thus far in this chapter. A key insight is that in order to exercise the full range of the quantizer output, the full VCO frequency range must also be exercised. As such, the nonlinearity of the volt-

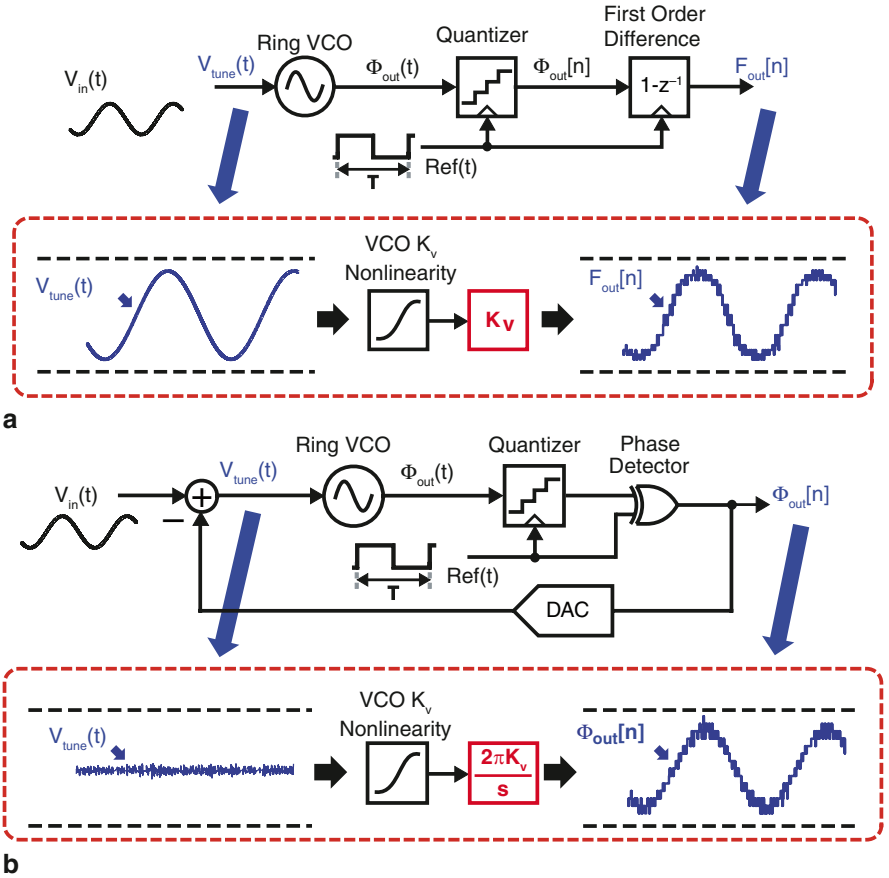


Fig. 11 Comparison of using Phase versus Frequency as the key variable of the VCO-based quantizer. **a** Frequency as the key VCO-based quantizer variable. **b** Phase as the key VCO-based quantizer variable

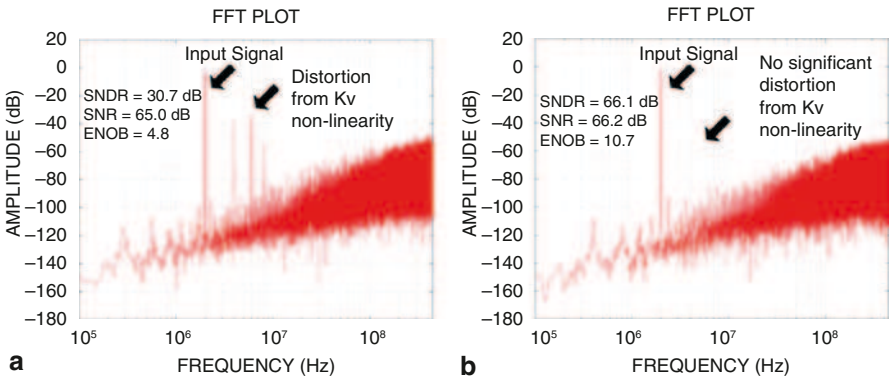


Fig. 12 Simulated spectra of systems in Fig. 11 which illustrate the dramatic reduction of harmonic distortion by using phase as the key variable of the VCO-based quantizer. **a** Frequency as key variable. **b** Phase as key variable

age-to-frequency characteristic is fully encountered which leads to a corresponding loss in SNDR performance.

In contrast, Fig. 11b displays the use of phase as the key quantizer variable, which is achieved by slight rearrangement of the XOR comparison circuit and the inclusion of DAC feedback. As illustrated in the figure, the use of phase as the key quantizer variable allows the full range of the quantizer output to be exercised with only slight deviation at the VCO input. The reduction in input deviation is made possible by the integrating action of the VCO in translating input voltage to output phase. Since only a small range of the voltage-to-frequency characteristic of the VCO is utilized, the relative impact of the nonlinearity of that characteristic is dramatically reduced compared to when frequency is the key variable. As such, the use of VCO phase opens the door to achieving higher SNDR in the overall ADC. We will see that only modest increase in ADC implementation complexity is required to achieve this improved SNDR.

To illustrate the benefits of using phase as the key variable of the VCO-based quantizer, Fig. 12 shows FFT plots from behavioral simulations of quantization noise occurring in the two systems shown in Fig. 11 [15]. SNDR is calculated assuming a 20 MHz bandwidth, a −1 dBFS input signal at 2 MHz, a sample rate of 1 GHz, and a 31-stage ring-VCO structure with *Kv* non-linearity modeled as a 4th order polynomial. As Fig. 12 reveals, using phase as the key variable suppresses harmonic distortion to the point where Kv non-linearity of the VCO is no longer the bottleneck to achieving high SNDR.

5 Striving for 80 dB SNDR with 20 MHz Bandwidth

Figure 12 reveals that from the standpoint of quantization noise, a first order structure such as shown in Fig. 11b is capable of SNDR values in the range of 66 dB with 20 MHz bandwidth. When seeking to improve SNDR values to close to the 80 dB

range, the quantization noise must be further reduced by increasing the oversampling ratio or the order of noise shaping in the Sigma-Delta ADC, or both. However, the simulation results from Fig. 12 already assume 1 GHz sample rate—a further increase in the sample rate (i.e., oversampling ratio) is challenging without going to a more advanced fabrication process than the 0.13 μ CMOS node assumed in the work presented here. Therefore, we now discuss issues related to increasing the order of the noise shaping while using VCO-based quantization with phase as the key variable.

As we examine the issue of achieving higher order noise shaping ADC topologies when using a VCO-based quantizer, it is worthwhile to first examine some tradeoffs that occur between using phase versus frequency as the key variable. We previously discussed the benefits of first order shaping of quantization noise and the intrinsic DEM operation offered by a VCO-based quantizer with frequency as the key output signal. Unfortunately, we lose both of these benefits when using phase as the key output signal due to the removal of the first order difference operation. However, using phase as the key output variable does have the advantage of providing an integrator with infinite gain at DC since phase is, in an ideal sense, the integral of frequency. An infinite DC gain integrator is quite useful within an ADC since it helps avoid dead zone issues [37, 38], and is especially relevant for more advanced CMOS processes which are facing steadily reduced intrinsic gain, i.e., $g_m r_o$, of CMOS devices.

The inclusion of an integrator within the quantizer must be considered when choosing the higher order ADC topology. As an example, Fig. 13a displays a higher order continuous-time Sigma-Delta ADC topology which utilizes a classical quantizer. The higher order dynamics are achieved by cascading integrators and combining their outputs using feed-forward paths. Also, the impact of quantizer delay on the stability of the ADC dynamics is mitigated with the use of an additional feedback DAC around the quantizer [2].

In contrast, Fig. 13b displays a corresponding topology based on a VCO-based quantizer with phase as its key output. One should notice that the primary difference between the two approaches is that the inherent integrator within the VCO-based quantizer allows one less explicit integrator to be implemented while still achieving the same order of noise shaping. However, the combined integrator/quantizer implementation forces the feedforward summing node to be placed before the integrator rather than the quantizer. As shown in Fig. 13b, this issue is addressed by replacing one of the feedforward paths with an additional DAC feedback path and modifying the feedforward coefficients accordingly [39]. Also, loop delay compensation is accomplished by differentiating the quantizer output (i.e., using frequency of the VCO-based quantizer) to compensate for the integrating action of the VCO-based quantizer. A side benefit of using frequency in the loop delay compensation feedback path is that its corresponding DAC will have its unit element mismatch noise shaped by the barrel shifting action of the VCO-based quantizer.

Implementation of the block diagram shown in Fig. 13b can be achieved in several different ways with respect to how the integrators and DACs are implemented. The two most popular choices of integrators are based on either opamps or gm-C

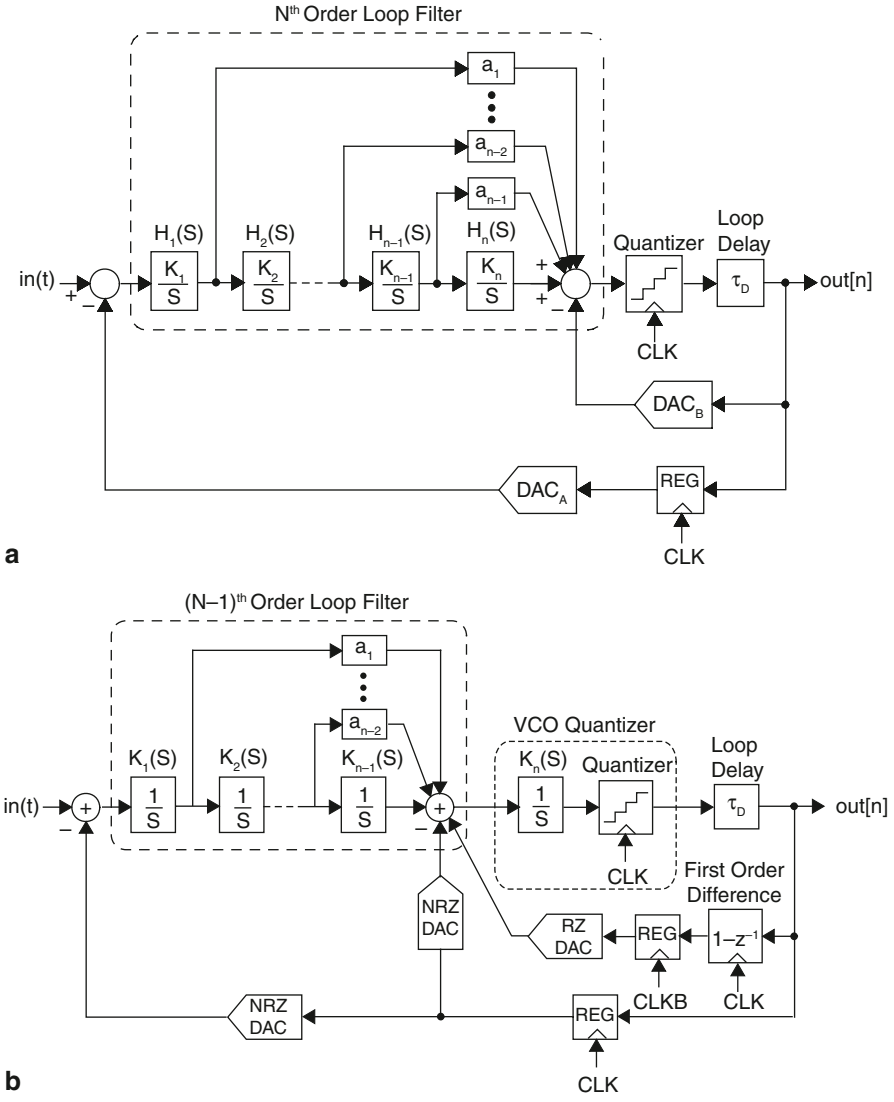


Fig. 13 Block diagram of a 4th order continuous-time Sigma-Delta ADC **a** using a traditional quantizer, **b** VCO-based quantizer with phase as key output variable

cells. As for DACs, they can be either current or switched capacitor based [40] and have a structure that utilizes Return-to-Zero (RZ) or Non-Return-to-Zero (NRZ) signaling.

Fig. 14 illustrates the ADC prototype considered in this chapter, which uses opamp-based integrators and current-based DACs. As shown in the figure, the DACs use a combination of RZ and NRZ signaling. Opamp-based integrators,

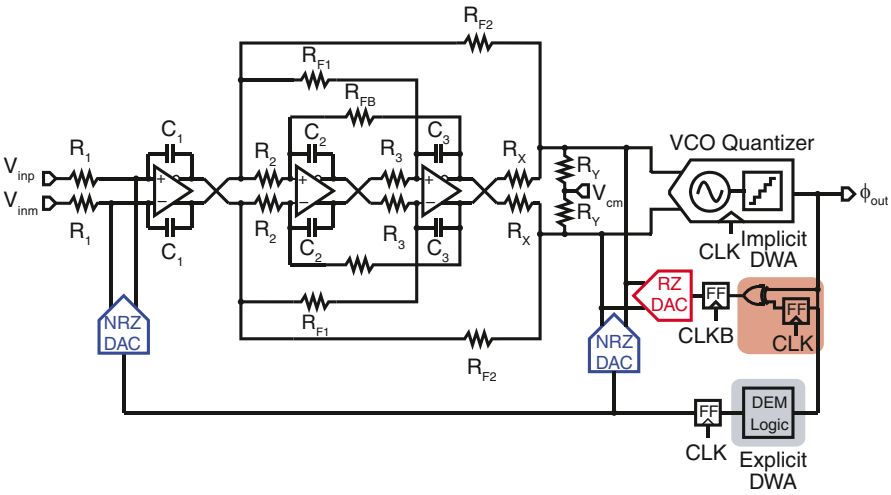


Fig. 14 Simplified circuit level implementation of the 4th order continuous-time Sigma-Delta ADC represented in Fig. 13

which are generally higher in power consumption compared to gm-C cells, are chosen for the advantages they offer in linearity performance. A RZ DAC is chosen in the minor loop around the quantizer in order to absorb the propagation delay of the quantizer [2]. However, for the main feedback DAC as well as an auxiliary minor loop feedback DAC, NRZ signaling is chosen in order to minimize the impact of the DAC switching waveform on the opamp inputs that they feed into [15]. For a continuous-time ADC, the concern is that large transients at opamp inputs will introduce distortion that can cause folding of the shaped quantization noise and therefore a reduction in SNR for the overall ADC.

In order to evaluate the potential performance of the ADC topology shown in Fig. 14, a detailed behavioral model was constructed in CppSim which is available for download [41]. In this model, key non-idealities such as VCO Kv nonlinearity, device noise, finite gain and finite bandwidth of the opamps, and mismatch between the DAC and VCO elements were included as estimated by SPICE simulations of the individual circuit blocks implemented in a 0.13 μ CMOS process. To counteract the issue of mismatch in the NRZ DAC elements, an explicit DEM circuit was included in the behavioral model as indicated in Fig. 14. However, note that the RZ DAC elements, which are controlled with frequency rather than phase as the key VCO-based quantizer variable, are intrinsically barrel shifted. As shown in Fig. 15, the behavioral simulation shows a theoretical peak SNDR performance of 85 dB with 20 MHz bandwidth with no harmonic distortion visible. As such, in theory, the VCO Kv nonlinearity estimated from a 0.13 μ VCO-based quantizer design does not present a bottleneck to achieved SNDR values in excess of 80 dB when using phase as the key variable for the VCO-based quantizer.

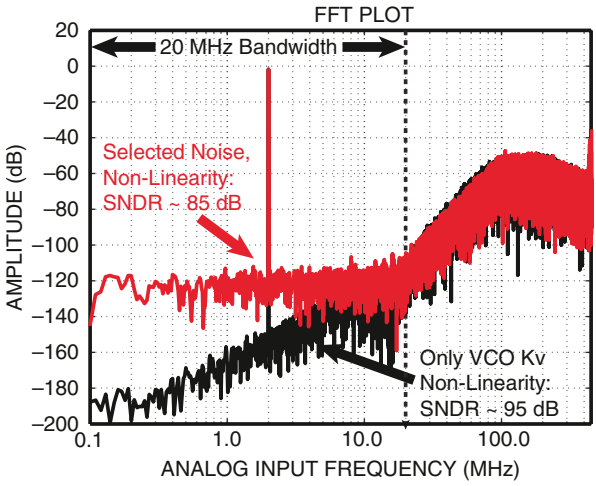


Fig. 15 Simulated output spectra of ADC topology shown in Fig. 14 based on detailed behavioral simulation with a 2 MHz input signal at −1 dBFS

6 Circuit Details of a 4th Order, Continuous-Time Sigma-Delta ADC Prototype in 0.13 µ CMOS

We now present circuit details of a 4th order, continuous-time Sigma-Delta ADC prototype that uses a VCO-based quantizer with phase as its key output variable. The system topology is based on Fig. 14, and here we will focus on the implementation of the VCO-based quantizer, integrator opamps, current DACs, and DEM circuit.

6.1 VCO-Based Quantizer

As shown in Fig. 16, the VCO-based quantizer consists of a ring oscillator voltage-controlled oscillator (VCO), a sense-amp flip-flop [42] for quantizing the VCO output phase, and additional XOR gates and registers to convert the quantized phase to frequency for the RZ feedback DAC shown in Fig. 14. Note that the quantizer used in the prototype is 4-bit, but a 3-bit version is shown in Fig. 16 for simplicity.

The ring oscillator VCO nominally oscillates at a frequency of 225 MHz, but can span from nearly 0 Hz to approximately 450 MHz when the control voltages are swept. The VCO delay element is based on a current starved inverter, and enables pseudo differential control as well as frequency and *Kv* tuning to cover process variations [14]. The phase detector reference signals also have a frequency of 225 MHz, and are generated by dividing the 900 MHz ADC clock frequency by 4. The quadrature relationship between Ref_0 and Ref_1 is obtained by using a register to delay one reference signal relative to the other by one sample period of the 900 MHz clock.

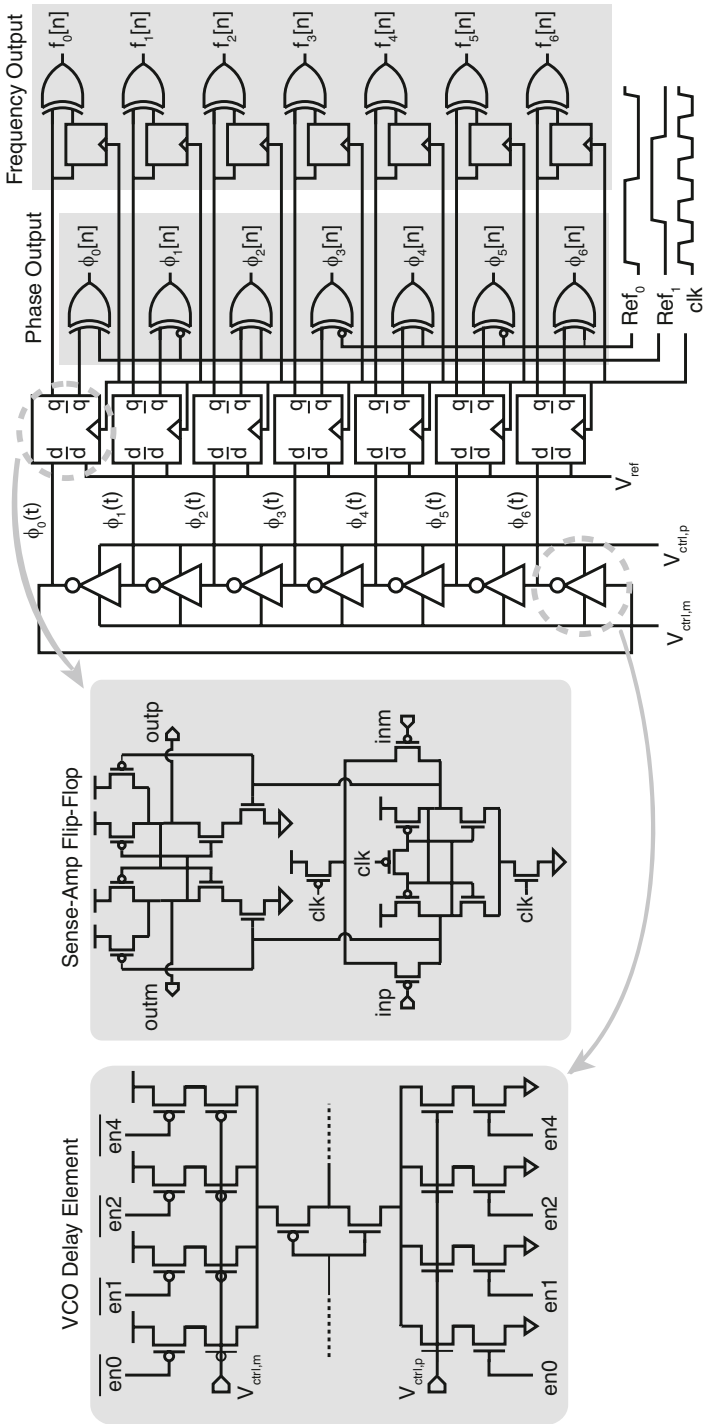


Fig. 16 Details of VCO-based quantizer implementation

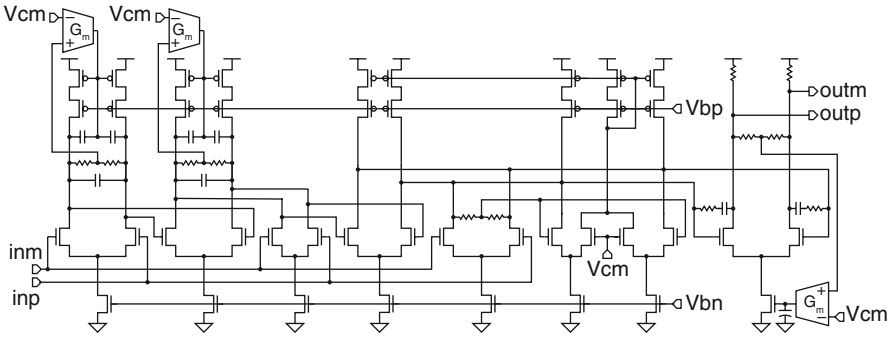


Fig. 17 Nested Miller Opamp Implementation [1, 43]

6.2 Opamp Design

Figure 17 shows the nested Miller topology used to realize the opamps shown in Fig. 14 [1, 43]. As shown in the figure, this opamp topology is composed of several cascaded gain stages in order to achieve a DC gain greater than 60 dB and unity gain bandwidth of 4 GHz. The high unity gain bandwidth is needed so that the opamp can sufficiently track the wideband output of the current DAC that feeds into its input in order to avoid excessive quantization noise folding. Note that many gain stages are needed in 0.13 μ CMOS due to the relatively low intrinsic gain (i.e., $g_m r_o$ product) offered by its devices. Stability is achieved by the inclusion of two feed-forward paths which introduce left-half plan zeros that compensate for the additional poles of the cascaded gain stages.

6.3 DAC Designs

As revealed by Fig. 14, two different current DAC topologies are used in the ADC prototype. Figure 18 shows the NRZ DAC implementation used in the main feedback path, which is the most critical DAC for achieving high SNDR in the overall ADC. As shown in the figure, the key element in the DAC is a differential pair whose current source is formed by a resistor degenerated cascode current mirror. The resistor degeneration lowers the impact of 1/f noise [38], and the cascode current mirror helps to prevent current variations as the output of the differential pair experiences voltage variations. Keeping constant current is important to minimize nonlinear distortion which could otherwise degrade the SNDR of the overall ADC. To further improve this issue, the inputs to the differential pair, *Vinp* and *Vinm*, are constrained to a small swing with the low-swing buffer shown in the figure. The small swing minimizes voltage variations experienced by the current source, and further boosts its output resistance by essentially adding an extra stage of cascod-

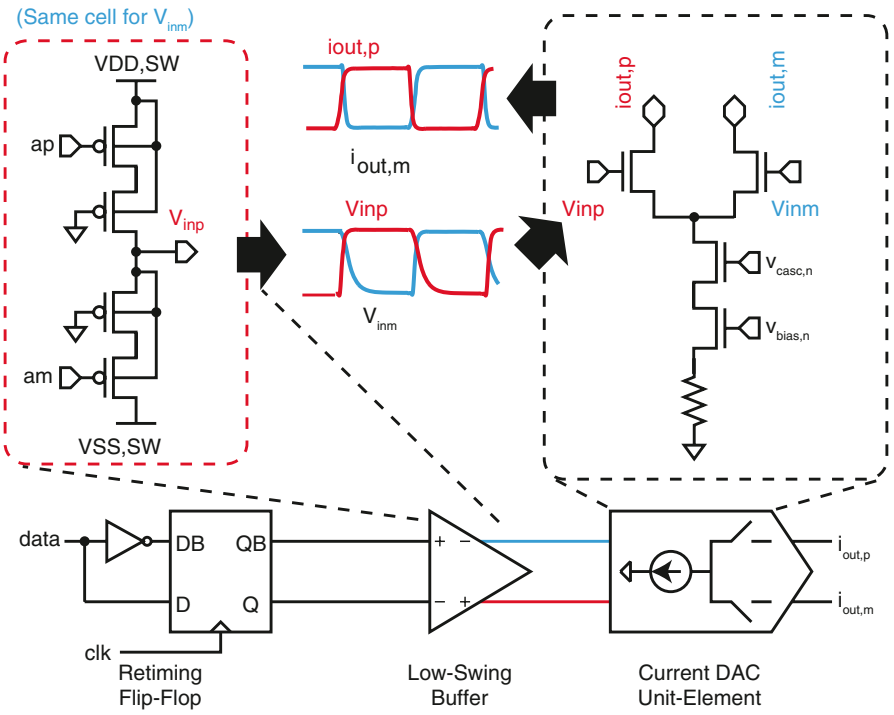


Fig. 18 Simplified schematic of the main feedback NRZ DAC

ing [15]. Finally, note that the other NRZ DAC (whose output connects to the RZ DAC output) has a similar implementation to that shown in Fig. 18, but is designed to consume less power and eliminates the degeneration resistor. These changes are acceptable since the DAC feeds into a node that is placed after several stages of gain in the ADC, so that its noise performance is not as critical for achieving high SNDR in the overall ADC.

Figure 19 shows a simplified schematic of the RZ DAC shown in Fig. 14 which helps compensate for quantizer latency in order to achieve stability in the overall ADC [13]. Since this DAC feeds into a node that is placed after several stages of gain in the ADC, its noise performance is not critical for achieving high SNDR in the overall ADC. As such, a simple implementation is in order, and we see from Fig. 19 that the prototype implements this DAC as three complementary current steering devices that are fed by full swing signals generated by digital logic gates. Note that the top PMOS current steering devices are controlled by the clock such that the top current feeds into nodes *outp* and *outm* while the clock is low and into a dump node while the clock is low. While the clock is low, the data signal controls whether the bottom current is sent to node *outp* or *outm*, thus creating either a positive or negative differential current. While the clock is high, the bottom current is

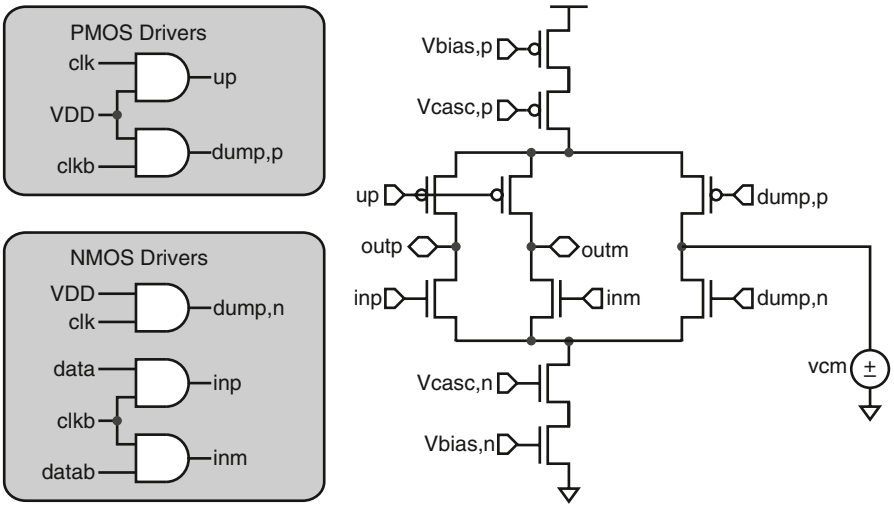


Fig. 19 Simplified schematic of the loop-delay compensating RZ DAC

fed into the dump node. Note that shifting the top and bottom currents into the dump node during the 'return to zero' portion of the DAC cycle allows faster transients when switching the current DAC back through the *outp* and *outm* nodes during the 'on' portion of the DAC cycle.

6.4 DEM Design

One disadvantage of using phase as the key output variable of the VCO-based quantizer is the loss of the intrinsic DEM operation offered by its frequency output counterpart. As such, we need to implement an explicit DEM as shown in Fig. 14. Figure 20 shows the topology chosen for this explicit DEM in the prototype, which provides a first order Dynamic Weighted Average (DWA) operation on the main feedback NRZ DAC [17]. As shown in the figure, the implemented DWA circuitry rotates the quantizer output, which is encoded in thermometer code format, with the aid of a barrel shifter controlled by a binary accumulator. Note that the number of times that the *current* quantizer thermometer code needs to be shifted by is equal to the modulo-2^N accumulated sum of the *previous* quantizer output values. Consequently, the DWA can be split into two parallel paths, one of which shifts the input thermometer code, and the other which records the previous quantizer value and updates the pointer. A fully static-CMOS implementation of the DWA in the 0.13 µm technology was not able to satisfy the timing requirements with sufficient margin due to the wiring parasitics and device capacitances. Instead, a faster pseudo

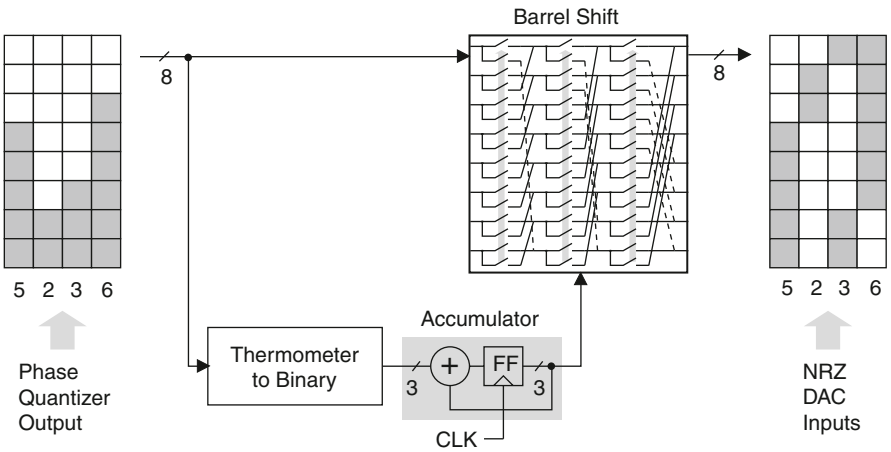


Fig. 20 Explicit DWA computation using a barrel shift, a thermometer-to-binary converter, and binary accumulator

NMOS logic implementation using PMOS loads was chosen [44], though at the cost of higher power consumption.

6.5 Overall Implementation in 0.13 μ CMOS

A die-photo of the fabricated prototype ADC in 0.13 μm CMOS is shown in Fig. 21. The active silicon area of the ADC is 0.45 mm^2, and the total chip area including 48 pads is 2.3 mm × 1.8 mm. The prototype ADC dissipates roughly 87 mW from a 1.5 V supply, with the analog and digital supplies drawing roughly 46 mA and 12 mA, respectively. Although there was no direct way to measure the subsystem current, bias currents indicate that the DACs consume 15 mA, the operational amplifiers consume 30 mA, and the VCO consumes less than 1 mA. Simulations indicate that the data-weighted averaging logic comprises the majority (>75%) of the digital power dissipation due to the use of pseudo NMOS logic in the thermometer-to-binary converter and accumulator, with VCO phase quantizer flip-flops and clock generation and distribution circuits comprising the remainder.

7 Measured Results of the Prototype

The test setup used to evaluate the 4th order prototype ADC is shown in Fig. 22. Here, an analog signal source (Agilent E4430B) drives a 2 MHz tone into a passive bandpass filter (TTE KC7T-2M-10P), which suppresses the harmonics and phase

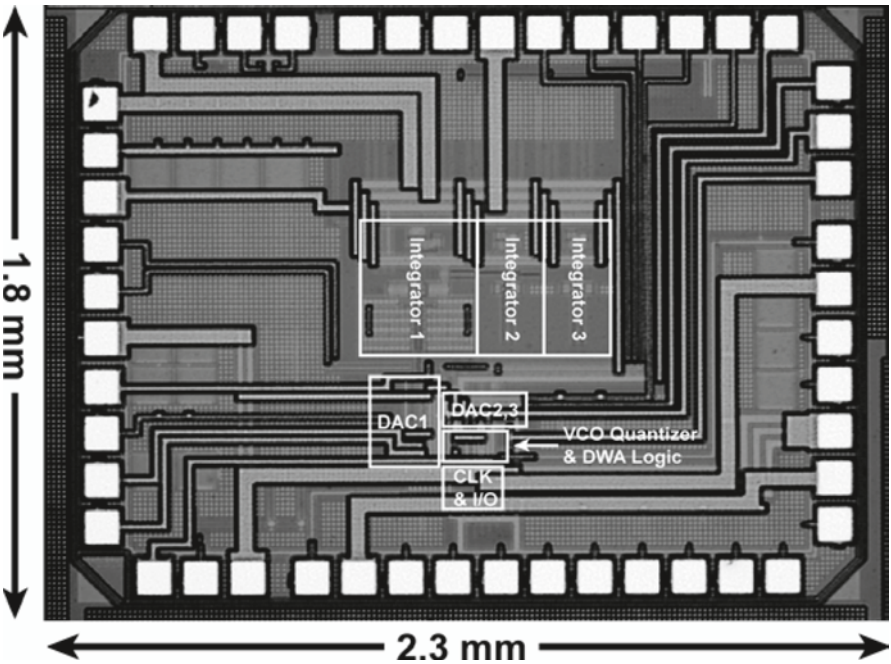


Fig. 21 Die photo of the prototype ADC fabricated in a 0.13 µm IBM CMOS process. The active area is 0.45 mm^2

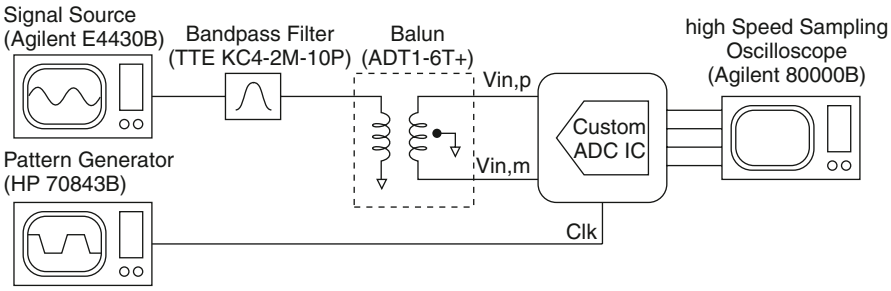


Fig. 22 Test setup for evaluation of the prototype ADC

noise of the signal source. An RF transformer (Mini-Circuits ADT1-6T+) converts this spectrally purified tone into a differential signal that serves as the input to the prototype ADC. The ADC clock signal is generated by a high-speed pattern generator (HP 70843B), which can generate low-jitter, square waveforms (<1 ps, RMS in bandwidth of interest). The 4 digital output bits generated by the prototype ADC are stored into the memory of a high-speed sampling oscilloscope (Agilent DSA 80000B), and then downloaded to a PC for post-processing.

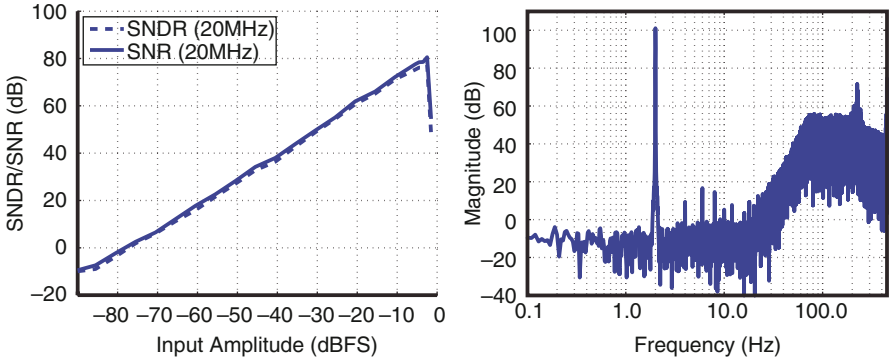


Fig. 23 Measured SNR/SNDR versus input amplitude, and 100,000 point FFT plot generated from measured output of prototype ADC

The measured SNR and SNDR versus input amplitude curves are shown in Fig. 23. For these measurements, the input tone frequency is 2 MHz, the analog bandwidth is 20 MHz, and the sample rate is 900 MHz. For a −2.4 dBFS (approximately 1.5 $V_{pp,diff}$) input tone, the ADC achieves a peak SNR of 81.2 dB, and a peak SNDR of 78.1 dB; this corresponds to a resolution of 12.7 ENOB, and a Figure of Merit (FOM) of 330 fJ/conversion, where:

$$FOM \equiv \frac{Power}{2 \times BW \times 2^{ENOB}}.$$

Figure 23 also shows a fast Fourier transform (FFT) of the ADC output for the 2 MHz input signal at −2.4 dBFS. Fourth-order quantization noise shaping is visible in the frequency range from 20 to 70 MHz, and peaks locally at 70 MHz due to a parasitic pole, which degrades phase margin. Behavioral simulations suggest that the tones appearing in the 200–250 MHz range and centered at 225 MHz (Fs/4) are most likely due to the phase detector reference clock signal and the VCO output phases parasitically coupling into the VCO control node. Fortunately, these tones are far out of band and did not affect the resolution or stability of the ADC.

Finally, Table 1 compares the prototype performance against recent continuous-time Sigma-Delta ADCs with similar input bandwidths and technology. As seen by this table, the use of phase as the key variable for VCO-based quantization provides competitive performance against more conventional ADC topologies. Since this prototype is the first pass at using VCO-based quantization in this manner, it is expected that future implementations will further improve SNDR according to the system level analysis results shown in Fig. 15.

Table 1 Comparison of recent CT Sigma-Delta ADCs with similar input bandwidths and technology

Ref.	Fs (MHz)	BW (MHz)	SNR (dB)	SNDR (dB)	Power (mW)	FOM (pJ/Conv)
[5]	276	23	70	69	46	0.43
[4]	340	20	71	69	56	0.61
[3]	400	12	64	61	70	3.18
[6]	640	10	72	66	7.5	0.23
[1]	640	20	76	74	20	0.12
[13]	640	10	87	82	100	0.49
[7]	1,000	8	63	63	10	0.54
[14]	950	20	75	67	40	0.55
This work [15]	*900*	*20*	*81*	*78*	*87*	*0.33*

8 Conclusions

This chapter examined the use of VCO-based quantization within wideband, continuous-time Sigma-Delta ADC topologies. We showed that a VCO-based quantizer can be viewed as an efficient combination of a Voltage-to-Time converter combined with a Time-to-Digital converter. While there are several advantages offered by this structure, including shaping of quantization noise and mismatch, this quantizer structure unfortunately has a nonlinear characteristic which can hamper the achievement of high SNDR values in the overall ADC. However, by using phase as the key output variable of the quantizer, we showed that this nonlinearity issue can be overcome such that over 80 dB of SNDR can be theoretically achieved with 20 MHz bandwidth. Measured results of a prototype ADC confirm 78 dB SNDR performance with 20 MHz bandwidth while achieving a 330 fJ/conversion step efficiency. As such, the VCO-based quantizer presents an attractive option for future high performance, continuous-time ADC applications.

References

1. G. Mitteregger, C. Ebner, S. Mechnig, T. Blon, C. Holuigue, E. Romani, A 20 mW 640-MHz CMOS continuous-time sigma-delta ADC with 20-MHz signal bandwidth, 80-dB dynamic range, and 12-bit ENOB. *IEEE J. Solid-State Circuits* **41**(12), 2641–2649 (2006)
2. S. Yan, E. Sanchez-Sinencio, A continuous-time modulator with 88-dB dynamic range and 1.1-MHz signal bandwidth. *IEEE J. Solid-State Circuits* **39**(1), 75–86 (2004)
3. S. Paton, A. Di Giandomenico, L. Hernandez, A. Wiesbauer, T. Potcher, M. Clara, A 70-mW 300-MHz CMOS continuous-time ΣΔ ADC with 15-MHz bandwidth and 11 bits of resolution. *IEEE J. Solid-State Circuits* **39**(7), 1056–1063 (2004)
4. L. Breems, R. Rutten, G. Wetzker, A cascaded continuous-time ΣΔ modulator with 67-dB dynamic range in 10 MHz bandwidth. *IEEE J. Solid-State Circuits* **39**(12), 2152–2160 (2004)
5. N. Yaghiniand, D. Johns, A 43 mW CT complex ΔΣ ADC with 23 MHz of signal bandwidth and 68.8 dB SNDR, in *IEEE ISSCC Dig. Tech. Papers*, 2005, pp. 502, 503

6. R. Schoofs, M. Steyaert, W. Sansen, A design-optimized continuous-time ΔΣ ADC for WLAN applications. *IEEE Trans. Circuits Syst. I, Reg. Papers* **54**(1), 209–217 (2007)
7. R. Schoofs, M. Steyaert, W. Sansen, A 1 GHz continuous-time sigma-delta A/D converter in 90 nm standard CMOS, in *IEEE MTTT-S Tech. Dig*, 2005, pp. 1287–1290
8. F. Esfahani, P. Basedau, R. Ryter, R. Becker, A fourth-order continuous-time complex sigma-delta ADC for low-IF GSM and EDGE receivers, in *IEEE Symp. on VLSI Circ. Dig. Tech. Papers*, 2003, pp. 75–78
9. L. Breems, R. Rutten, R. Veldhoven, G. Weide, H. Termeer, A 56 mW CT quadrature cascaded ΣΔ modulator with 77 dB DR in a near zero-IF 20 MHz band, in *IEEE ISSCC Dig. Tech. Papers*, 2007, pp. 238–599
10. R. Veldhoven, R. Rutten, L. Breems, An inverter-based hybrid ΣΔ modulator, in *IEEE ISSCC Dig. Tech. Papers*, 2008, pp. 492–630
11. Y. Shu, B. Song, K. Bacrania, A 65 nm CMOS CT ΣΔ Modulator with 81 dB DR and 8 MHz BW Auto-Tuned by Pulse Injection, in *IEEE ISSCC Dig. Tech. Papers*, 2008, pp. 500–631
12. S. Ouzounov, R. Veldhoven, C. Bastiaansen, K. Vongehr, R. Wegberg, G. Geelen, L. Breems, A. Roermund, A 1.2 V 121-mode CT ΣΔ modulator for wireless receivers in 90 nm CMOS, in *IEEE ISSCC Dig. Tech. Papers*, 2008, pp. 242–600
13. W. Yang, W. Schofield, H. Shibata, S. Korrapati, A. Shaikh, N. Abaskharoun, D. Ribner, A 100 mW 10 MHz-BW CT ΔΣ modulator with 87 dB DR and 91 dBc IMD, in *IEEE ISSCC Dig. Tech. Papers*, 2008, pp. 498–631
14. M. Straayer, M. Perrott, A 12-bit 10-MHz bandwidth, continuous-time sigma-delta ADC with a 5-bit, 950-MS/S VCO-based quantizer. *IEEE J. Solid-State Circuits* **43**(Apr.), 805–814 (2008)
15. M. Park, M.H. Perrott, A 78 dB SNDR 87 mW 20 MHz bandwidth continuous-time delta-sigma ADC with VCO-based integrator and quantizer implemented in 0.13 μm CMOS. *IEEE J. Solid-State Circuits* **44**(Dec.), 3344–3358 (2009)
16. S.-J. Huang, Y.-Y. Lin, A 1.2 V 2 MHz BW 0.084 mm^2 CT ΔΣ ADC with −97.7 dBc THD and 80 dB DR using low-latency DEM, in *IEEE ISSCC Dig. Tech. Papers*, 2009, pp. 172, 173
17. R. Baird, T. Fiez, Linearity enhancement of multibit delta-sigma A/D and D/A converters using data weighted averaging. *IEEE Trans. Circuits Syst. II, Analog Digit. Signal Process.* **42**(7), 753–762 (1995)
18. U. Wismar, D. Wisland, P. Andreani, A 0.2 V 0.44 μW, 20 kHz ΣΔ analog to digital modulator with 57 fJ/conversion FoM, in *Proceedings of IEEE European Solid State Circuits Conference*, 2006, pp. 187–190
19. U. Wismar, D. Wisland, P. Andreani, A 0.2 V 7.5 μW, 20 kHz ΣΔ modulator with 69 dB SNR in 90 nm CMOS, in *Proceedings of IEEE European Solid State Circuits Conference*, 2007, pp. 206–209
20. A. Iwata, N. Sakimura, M. Nagata, T. Morie, The architecture of ΔΣ analog-to-digital converters using a voltage-controlled oscillator as a multibit quantizer. *IEEE. Trans. Circuits Syst. II* **46**(7), 941–945 (1999)
21. R. Naiknaware, H. Tang, T.S. Fiez, Time-referenced single-path multi-bit ADC using a VCO-based quantizer. *IEEE Trans. Circuits Syst. II* **47**(7), 596–602 (2000)
22. J. Kim, S. Cho, A time-based analog-to-digital converter using a multi-phase voltage controlled oscillator, in *Proceedings of the 2006 IEEE Int. Symp. on Circuits and Systems*, May 2006, pp. 3934–3937
23. V.B. Boros, A digital proportional integral and derivative feedback controller for power conditioning equipment, in *IEEE Power Electronics Specialists Conf. Rec.*, June 1977, pp. 135–141
24. J.P. Hurrell, D.C. Pridmore-Brown, A.H. Silver, Analog-to-digital conversion with unlatched SQUID's. *IEEE Trans. Electron Devices* **ED-27**(10), 1887–1896 (1980)
25. E. Alon, V. Stojanovic, M.A. Horowitz, Circuits and techniques for high-resolution measurement of on-chip power supply noise. *IEEE J. Solid-State Circuits* **40**(4), 820–828 (2005)
26. M. Park, M.H. Perrott, A single-slope 80 MS/s ADC using two-step time-to-digital conversion, in *ISCAS 2009*, May 2009, pp. 1125–1128

27. S. Naraghi, M. Courcy, M.P. Flynn, A 9 bit 14 μW 0.06 mm^2 pulse position modulation ADC in 90 nm digital CMOS, in *IEEE International Solid State Circuits Conference (ISSCC)*, Feb. 2009, pp. 168, 169
28. R. Staszewski, S. Vemulapalli, P. Vallur, J. Wallberg, P.T. Balsara, 1.3 V, 20 ps time-to-digital converter for frequency synthesis in 90-nm CMOS. *IEEE Trans. Circuits Syst. II* **53**(3), 2240–2244 (2006)
29. S. Henzler, S. Koeppe, W. Kamp, H. Mulatz, D. Schmitt-Landsiedel, 90 nm 4.7 ps-resolution 0.7-LSB single-shot precision and 19 pJ-per-shot local passive interpolation time-to-digital converter with on-chip characterization, in *IEEE ISSCC Dig. Tech. Papers*, 2008, pp. 548, 549
30. M. Lee, A. Abidi, A 9b, 1.25 ps resolution coarse-fine time-to-digital converter in 90 nm CMOS that amplifies a time residue. *IEEE J. Solid-State Circuits* **43**(4), 769–777 (2008)
31. V. Ramakrishnan, P.T. Balsara, A wide-range, high-resolution, compact, CMOS time to digital converter, in *VLSI Design (VLSID'06)*, Jan. 2006
32. J.-P. Jansson, A. Mantyniemi, J. Kostamovaara, A CMOS time-to-digital converter with better than 10 ps single-shot precision. *IEEE J. Solid-State Circuits* **41**(6), 1286–1296 (2006)
33. M.Z. Straayer, M.H. Perrott, A multi-path gated ring oscillator TDC with first-order noise shaping. *IEEE J. Solid-State Circuits* **44**(Apr.), 1089–1098 (2009)
34. M. Hovin, A. Olsen, T. Sverre, C. Toumazou, ΔΣ modulators using frequency-modulated intermediate values. *IEEE. J. Solid-State Circuits* **32**(1), 13–22 (1997)
35. M. Miller, Multi-bit continuous-time sigma-delta ADC, U.S. Patent 6,700,520, Mar. 2, 2004
36. G. Taylor, I. Galton, A mostly digital variable-rate continuous-time ADC delta-sigma modulator, in *IEEE International Solid-State Circuits Conference*, Feb. 2010, pp. 298, 299
37. P.J.A. Naus, E.C. Dijkmans, Low signal-level distortion in sigma delta modulators, 84th Convention of the Audio Engineering Society. *J. Audio Eng. Soc.* **36**(May), 382(1988)
38. S. Norsworthy, R. Schreier, G. Temes, *Delta-sigma data converters*, (Wiley, New York, 1996)
39. R. Schreier, ΔΣ Toolbox (MATLAB Central, 2009), http://www.mathworks.com/matlabcentral/fileexchange/19-delta-sigma-toolbox (The toolbox is generated in 2000 and updated in 2009)
40. R.H.M van Veldhoven, A triple mode continuous-time ΣΔ modulator with switched-capacitor feedback DAC for GSM EDGE/UMTS/CDMA2000 receiver. *IEEE. J. Solid-state Circuits* **38**(12), 2069–2076 (2003)
41. M. Park, Behavioral Simulation of a 4th-Order CT Delta-Sigma ADC with VCO-Based Integrator and Quantizer. (CppSim System Simulator, 2008), http://www.cppsim.com/tutorials.html
42. B. Nikolic, V.G. Oklobzija, V. Stojanovic, W. Jia, J.K.-S. Chiu, M.M.-T. Leung, Improved sense amplifier-based flip-flop: Design and measurements. *IEEE J. Solid-State Circuits* **35**(6), 876–884 (2000)
43. X. Fan, C. Mishra, E. Sanchez-Sinencio, Single miller capacitor frequency compensation technique for low-power multistage amplifiers. *IEEE J. Solid-State Circuits* **40**(3), 584–592 (2005)
44. J. Rabaey, *Digital integrated circuits: A design perspective* (Prentice-Hall, Upper Saddle River, 1996)

Wideband Continuous-Time Multi-Bit Delta-Sigma ADCs

J. Silva-Martinez, C.-Y. Lu, M. Onabajo, F. Silva-Rivas, V. Dhanasekaran and M. Gambhir

1 Introduction

Recent developments in mobile computing and wireless internet have led to exponential growth in demand for portable computers and smart phones equipped with WLAN operating at 802.11 g/n standards. The digital computing required by these gadgets is facilitated by process scaling that follows Moore's law and is expected to continue down to 10 nm physical gate lengths. However, integration of effective high-performance analog-to-digital converters (ADCs) in these technologies remains a challenge.

The delta-sigma ($\Delta\Sigma$) topology is an effective ADC architecture because a substantial part of the signal processing is performed in the digital domain. The analog-to-digital conversion is performed with a few analog blocks. Figure 1a shows a block diagram of a 1-bit $\Delta\Sigma$ modulator. By coupling efficient digital signal processing operations with analog circuits having minimal or no matching requirements, the 1-bit $\Delta\Sigma$ architecture is a promising candidate for nanometric technologies. Component matching is not a major concern since the 1-bit DAC is inherently linear. Nonetheless, application of a 1-bit $\Delta\Sigma$ modulator as ADC for WLAN faces several obstacles. A high over-sampling ratio (OSR) is needed to meet the signal-to-quantization noise ratio (SQNR) requirements. This leads to increased power dissipation in integrators as well as the decimation filter. On the other hand, achieving the required SQNR with low OSR and higher order filtering is limited by overloading effects and reduced stability of the modulator [1, 2]. Contrary to the 1-bit DAC, the 1-bit quantizer is quite nonlinear and its quantization noise is excessive when used in high-performance applications.

One of the alternatives to improve the SQNR without increasing the switching frequency is to use a multi-level quantizer and a multi-bit feedback DAC as illustrated in Fig. 1b. With this approach, the noise-shaping gain required in the loop filter can be relaxed as a result of the smaller quantization noise associated

J. Silva-Martinez (✉)
Electrical and Computer Engineering, Analog and Mixed-Signal Center,
Texas A&M University, College Station, TX, USA

H. Casier et al. (eds.), *Analog Circuit Design,*
DOI 10.1007/978-94-007-0391-9_11, © Springer Science+Business Media B.V. 2011

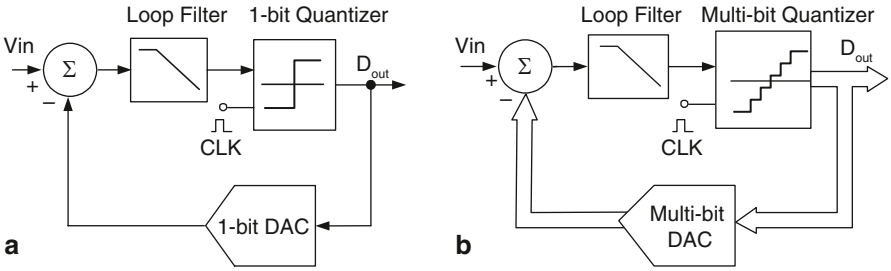


Fig. 1 **a** 1-bit ΔΣ modulator and **b** multi-bit ΔΣ modulator

with the multi-level quantizer. Multi-bit architectures have been successfully used in multi-MHz bandwidth designs [3–8]. However, the "digital friendly" advantages of the 1-bit architecture are typically compromised in such solutions. Specifically, the nonlinearity of the feedback DAC significantly affects the ADC performance since it directly adds error to the filter input signal that is not noise-shaped. As a result, dynamic element matching (DEM) techniques have been proposed to tackle this problem [9, 10]. However, the shaping of mismatch error provided by these techniques is inadequate in case of low OSR designs and less practical with high conversion speeds. Furthermore, the additional power and complexity of DEM methods is not permissible in some applications. A couple of solutions are discussed in Sects. 2 and 3.

2 A 20 MHz 68 dB Dynamic Range ΔΣ ADC Based on Time-Domain Quantizer and Feedback Element

One of the main challenges in implementing high-speed and high-precision analog functions in modern digital CMOS technologies is the low supply voltage. An important consequence of using low supply voltage is the reduced signal voltage swings. This in turn demands smaller noise levels for a given signal to noise ratio (SNR). To mitigate loss of dynamic range while scaling voltages, time resolution of discrete-amplitude waveforms can be used to represent amplitude variations of signals with an equivalent discrete-time duration and fixed amplitude. Consequently, the dynamic range depends on fine time resolution, and the highest switching frequency in the circuits is a function of the input signal frequency. This creates a dependence of the power dissipation on signal activity, which in many cases helps to reduce the average power consumption [11]. Since the signal representation is in the time domain, it is more critical to minimize the noise induced by clock jitter as compared to conventional multi-current sources DAC where process variations and thermal noise in the feedback path limit the performance of the ADC. The focus in this section is to give an overview of a time-resolution based ADC architecture that achieves more than 10-bit resolution over a 20 MHz bandwidth, which was designed in TI 65 nm digital CMOS technology.

2.1 Multi-Bit Time-Based ADC Architecture [12, 13]

A forward-looking strategy for architectural ΔΣ modulator improvements is to (a) use digital circuits wherever possible to take full advantage of the CV^2f rule for power reduction, and (b) represent signals in the time domain to leverage the fine time resolution available in the scaled technologies while avoiding the limitations associated with the lack of voltage headroom as well as rising process variations. The ΔΣ architecture based on a time-domain quantizer/DAC combination is shown in Fig. 2; where the pulse width modulation (PWM) generator encodes and samples incoming signals to continuous-in-time single-leveled signals, and the time-to-digital converter (TDC) interfaced to a single-level DAC replaces the conventional multi-level quantizer and the N-bit DAC.

During each clock period, a pulse with discrete levels of width represents the sample. The PWM generator block consists of a track-and-hold amplifier (THA) followed by a comparator that is synchronized with the master clock. This building block provides a pulse that has a width proportional to the amplitude of its input signal during every clock period. The TDC block digitizes the timing information into a pair of digital codes corresponding to the falling and rising instants. It also generates a "time-quantized" feedback pulse that has discrete levels in time steps of T_Q for constructing the DAC feedback signal. A differential pair is used for the generation of the DAC current pulses in order to achieve good power supply rejection with a reasonably accurate reference.

2.2 Lowpass Filter Design

The loop filter is designed to achieve 3rd-order quasi-inverse-Chebyshev high-pass noise shaping. Let H(z) represent the discrete-time (DT) filter that yields such noise transfer function:

$$H(z) = \frac{1.622z^2 - 2.093z + 0.8024}{z^3 - 2.922z^2 + 2.894z - 0.9721}. \quad (1)$$

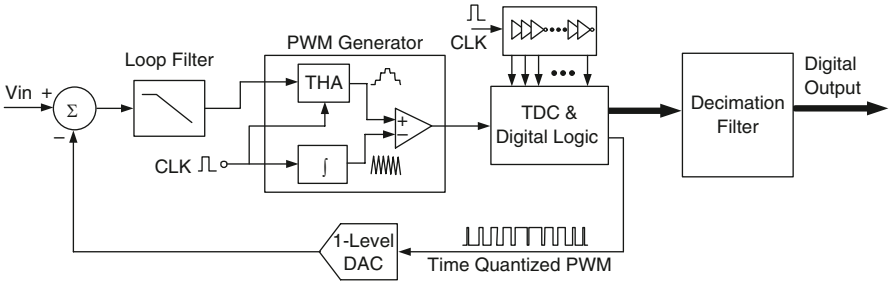


Fig. 2 Time-domain quantizer/DAC based delta-sigma modulator

The combined delay from the PWM comparator and the digital logic used to generate the feedback pulse was computed to be 660 ps. This is roughly equal to one-third of the 2 ns clock period. The loop delay due to the digital logic can be effectively compensated by realizing $H_2(z^{1/3})=z^{1/3}\cdot H(z^{1/3})$. The 3rd-order transfer function of the resulting DT filter is given by

$$H(Z^{1/3}) = \frac{0.4399Z^{2/3} - 0.7751Z^{1/3} + 0.3477}{Z^{3/3} - 2.985Z^{2/3} + 2.9764Z^{1/3} - 0.9906}. \tag{2}$$

Taking the $z^{1/3}$ delay introduced by the quantizer and DAC into account, and adding a constant term, we arrive at the following form:

$$\begin{aligned} Z^{1/3}H\left(Z^{1/3}\right) &= 0.4399 + H_2\left(Z^{1/3}\right) \\ &= 0.4399 + \frac{0.538Z^{2/3} - 0.9613Z^{1/3} + 0.4358}{Z^{3/3} - 2.985Z^{2/3} + 2.9764Z^{1/3} - 0.9906}, \end{aligned} \tag{3}$$

where the constant term represents the coefficient of a feedback path around the quantizer that compensates for the excess loop delay [14]. $H_2(z^{1/3})$ can be transformed to the equivalent CT transfer function by employing the impulse invariant $Z \rightarrow S$ transformation, leading to

$$H_2(s) = \frac{7.312s^2 + 2.312 \times 10^{17}7s + 4.223 \times 10^{25}}{s(s^2 + 1.414 \times 10^7 s + 1.279 \times 10^{16})}. \tag{4}$$

$H_2(s)$ provides a finite in-band gain of 37 dB, which serves to suppress quantization noise and other errors introduced by the subsequent blocks in the forward path. The simplified single-ended version of the active-RC filter topology realizing the transfer function is shown in Fig. 3, in which the feedforward capacitances C_B

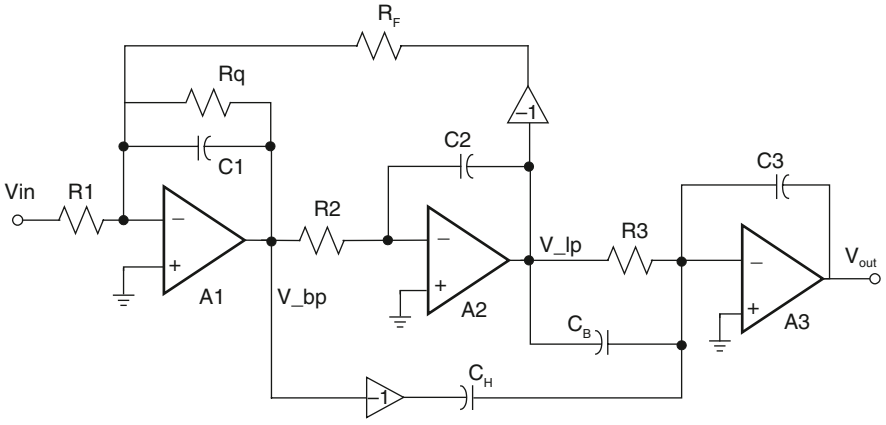


Fig. 3 Active-RC loop filter with feed-forward trajectories

and C_H provide the second-order bandpass and high-pass outputs from the biquads, respectively.

2.3 Time-Domain Quantizer

The pulse-width modulated (PWM) waveform is generated by comparing the input signal with a triangular waveform. The generator consists of a sample-and-hold (S/H) circuit, a ramp generator, and a pair of comparators. The PWM generator in Fig. 2 uses a ramp waveform at sampling frequency with differential amplitude of 1.2 V_{pp}. Although the data throughput from the time-domain quantizer is 250 MSPS, the S/H before quantization is clocked at 500 MHz (double-sampled PWM). This allows to double the OSR for a given throughput rate. The PWM-to-digital conversion is accomplished using a time-to-digital converter (TDC), which was initially proposed to measure single-shot pulses in nuclear experiments [15].

The functionality of the TDC implementation is illustrated in Fig. 4, for which details can found in [13]. There are N (=8 in this example) equally spaced time

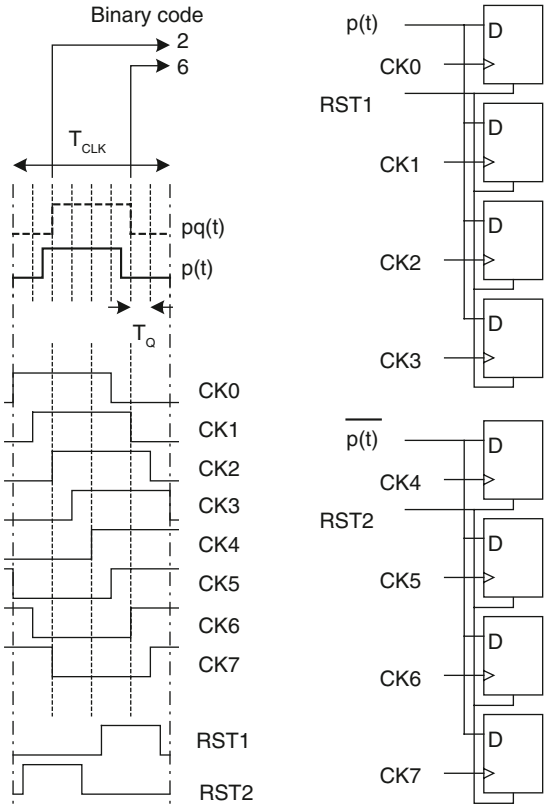


Fig. 4 Output code generation circuit

steps of length T_Q within a master clock period (T_{CLK}). The THA, ramp generator, and comparator create a time window p(t). This time frame is quantized using the clock phases CK0…CK7 to trigger a series of D flip-flops. The TDC block provides two output codes that correspond to the rising edge and the falling edge of the input pulse coming from the PWM generator. The time-quantized feedback pulse, denoted as pq(t), changes at the same time as the clock transitions (rising edge of CK0–7). The time-quantized feedback pulse, pq(t), is generated with a pair of OR gates and a SR latch. The output of the D flip-flops is decoded from thermometric code to binary code in order to generate the desired output. In a double-sampled PWM signal, the information is contained in the pulse width as well as the pulse position. To avoid aliasing of PWM tones, a pulse-to-amplitude converter is used to extract the final digital output.

The actual TDC was designed to generate 50 quantization steps in a 4 ns period (1/250 MHz). To support this scheme, the 50 clock phases are generated using digital inverter delay elements in cascade as represented in Fig. 2. Each of the clock phases generated with these delay elements is used to drive the clock input of a flip-flop. Robust clock generation can be accomplished by employing a phase-locked loop or a ring oscillator healed through injection-locking techniques [16]. The data-dependent delay or "metastability" of the flip-flops can lead to signal-dependent error in the feedback pulse. Since the maximum error corresponds to one quantization step (1/50th of full scale), the probability of occurrence of this error has to be kept at 1.6% or below to meet a distortion performance better than −70 dB. Also, the static mismatch between the quantization steps of the TDC contributes to nonlinearity in the feedback. Based on the simulations of the ADC model, it was determined that the RMS mismatch between the time steps of the delay elements of the TDC must be less than 800 fs to achieve a distortion performance better than −65 dB.

2.4 Spectrum of the PWM Signal DAC

The PWM spectrum consists of tones at the signal frequency (ω_S) and its odd harmonics, reference tones due to the ramp's fundamental frequency (ω_R) and its harmonics, as well as intermodulation products of the signal and the reference tones. The spectral content of a double-sampled PWM waveform can be shown to be [17]

$$
\begin{aligned}
v_p(t) = {} & \frac{2V_a}{\pi}\sum_{n=1}^{\infty}\frac{1}{\left[n\frac{\omega_S}{\omega_R}\right]}J_n\left(n\frac{\omega_S}{\omega_R}\frac{\pi}{2}M\right)Sin\left(n\frac{\pi}{2}\right)Cos\,(n\omega_S t) \\
& + \frac{2V_a}{\pi}\sum_{m=1}^{\infty}\frac{1}{m}J_o\left(m\frac{\pi}{2}M\right)Sin\left(m\frac{\pi}{2}\right)Cos\,(m\omega_R t) \\
& + \frac{2V_a}{\pi}\sum_{m=1}^{\infty}\sum_{n=-\infty}^{\infty}\frac{J_n\left(\left[m+n\frac{\omega_S}{\omega_R}\right]\frac{\pi}{2}M\right)}{\left[m+n\frac{\omega_S}{\omega_R}\right]}Sin\left([m+n]\frac{\pi}{2}\right)Cos\,(m\omega_R t + n\omega_S t).
\end{aligned}
\tag{5}
$$

Here, V_a represents the PWM signal amplitude, J_n the Bessel function of order n, and M the modulation index of the PWM. A prominent power content of the PWM spectrum is carried by higher order reference harmonics. Since, the TDC quantizes the PWM waveform in the time domain, the feedback pulse generated by the TDC is essentially a sampled and held version of the PWM pulse with a sampling period of T_Q. This sampling process inevitably results in aliasing of the high-frequency tones present in the spectrum of the input PWM waveform. It can be observed that a "quantization noise floor" is formed due to the TDC sampling. This is similar to the quantization noise floor in the conventional sampled amplitude quantizer, where the harmonic distortion of the quantized signal folds over to form a quantization noise floor due to sampling [11]. For a given number of "quantization steps", the time-domain quantizer is expected to have a different SQNR compared to the amplitude quantizer.

Due to the complexity of (5), it is more constructive to assess the quantization noise through simulations. The SQNR vs. number of quantization steps (N_Q) for amplitude and time-based quantizers can be obtained by considering the "integrated noise" over the frequency range from 0 to $1/(2T_s)$. It can be shown that the time quantizer exhibits higher quantization noise (due to aliasing of PWM intermodulation components rather than just the signal harmonics) for a given N_Q. This difference narrows for large N_Q (>50) due to amplitude flattening of the intermodulation components at very high frequencies, which reduces the difference to about 8 dB for number of quantization steps between 50 and 200 [13]. It is important to point out that higher quantization noise is not necessarily a severe performance disadvantage because the time quantizer can efficiently implement a larger number of quantization steps compared to the amplitude quantizer.

2.5 Design Considerations

There are various noise contributors in the time-based ADC architecture under investigation. The input-referred in-band RMS voltage noise of the filter (v_{nLF}) directly appears at the ADC input and thus contributes directly to the input-referred noise. The DAC and the resistor at the input (R_1) inject noise current into the virtual ground of the first integrator's operational amplifier in Fig. 3. Consequently, the in-band RMS current noise at the DAC output (i_{nDAC}) translates to input-referred voltage noise through a multiplicative factor determined by R_1. The output-referred in-band RMS timing jitter of the TDC (t_{nTDC}) can be mapped to the input-referred noise by appropriately scaling it with T_s, I_{ref}, and R_1; where T_s is the clock period (4 ns) and I_{ref} is the reference current of the DAC. The input-referred noise contribution due to the input-referred RMS voltage noise of the PWM generator (v_{nPWM}) is simply $v_{nPWM}/|H_{LF}|$, where $|H_{LF}|$ is defined as the average in-band gain of the loop filter over the 20 MHz signal bandwidth. Finally, the overall input-referred noise can be expressed as

Table 1 Noise contribution of various blocks within the time-based ADC

Block	Integrated input-referred voltage noise (μV_{RMS})	Percentage noise power contribution
Loop filter	85	41.6
DAC	66	25.1
TDC	67	25.8
PWM	36	7.5

$$v_{nin} = \sqrt{v_{nLF}{}^2 + i_{nDAC}{}^2 \cdot R_1{}^2 + \frac{t_{nTDC}{}^2}{T_s{}^2} \cdot Iref^2 \cdot R_1{}^2 + \frac{v_{nPWM}{}^2}{|H_{LF}|^2}}. \quad (6)$$

In this design, the differential reference voltage of the ADC was fixed at 1.08 V, the input resistance of the loop filter (R_1) was chosen as 3 kΩ, and the DAC reference current was set to 180 μA. The loop filter has an integrated input-referred RMS noise of 84.5 μV. Table 1 lists a breakup of the noise power contributions from various blocks in this example design.

The static mismatch between the quantization steps of the TDC contributes to nonlinearity in the feedback. Based on SIMULINK simulations of the ADC model, it was determined that the RMS mismatch between the time steps of the delay elements in the TDC must be less than 800 fs to achieve a distortion performance better than −65 dB with a 95% confidence level.

Another important design constraint is loop stability in presence of excess phase and delay. The main contributions to the excess delay are: propagation delay in the digital logic of the TDC that is used to generate the feedback pulse, the PWM comparator's delay and the excess phase of the loop filter. One option to preserve the NTF is by redesigning the loop filter such that it has a transfer function $H(z) \cdot z^{+\Delta}$ instead of the desired filter transfer function H(z). Here, Δ denotes the fractional excess delay (the ratio of excess delay-time to the sampling period) in the loop. This redesign requires an additional feedback path around the quantizer to ensure controllability of the system [14]. Figure 5 displays the arrangement of the compensation method. A bank of binary weighted CMOS inverters and MOS capacitors are used to facilitate programming of the feedback coefficient.

Clock jitter remains as one of the performance-limiting parameters for a continuous-time modulator. The error introduced in the DAC pulse attributable to the clock jitter is not noise-shaped by the loop filter because it directly appears at the modulator's input. This error can be modeled as a random phase modulation of the feedback pulses. In this case, some improvement in jitter performance is achieved due to the particular feedback pulse arrangement. From Fig. 6, it can be seen that both the rising and the trailing edge of the feedback pulse within a clock period experience almost the same time shift due to the low and medium frequency clock jitter; the main effect of the clock jitter is reflected in the position of the pulse. This is caused by the fact that rising edges of CK0–7 are derived from the same clock edge using delay elements. Since the jitter-induced time shift of both clock edges is almost identical, the amount of feedback charge returned by the PWM remains

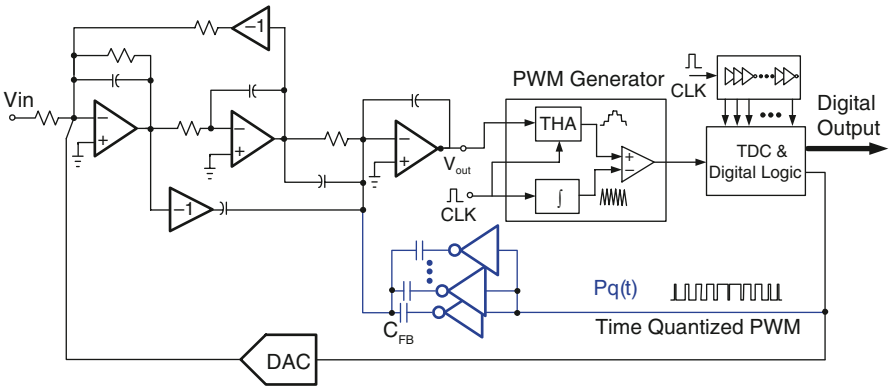


Fig. 5 Simplified ADC architecture

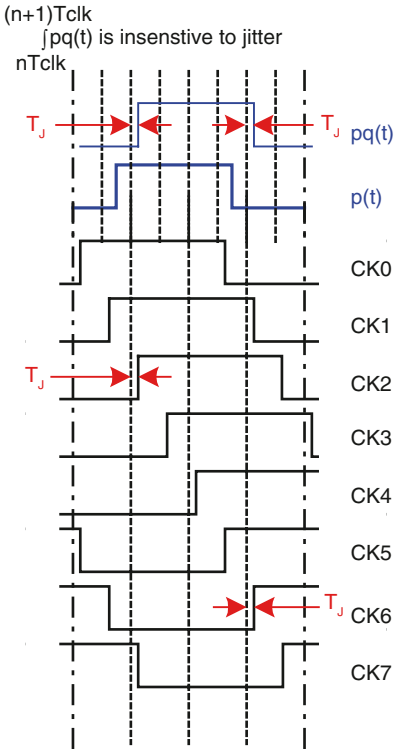


Fig. 6 Effect of clock jitter in the modulator with PWM feedback

the same. Hence, the jitter from the clock source mainly affects the position of the feedback pulse; leaving only the jitter noise added by each individual delay element as main jitter noise contributor to feedback charge injection error.

The effect of random pulse position modulation of a PWM waveform is cumbersome to analyze with equations, but the fundamental result will be given in Sect. 3.

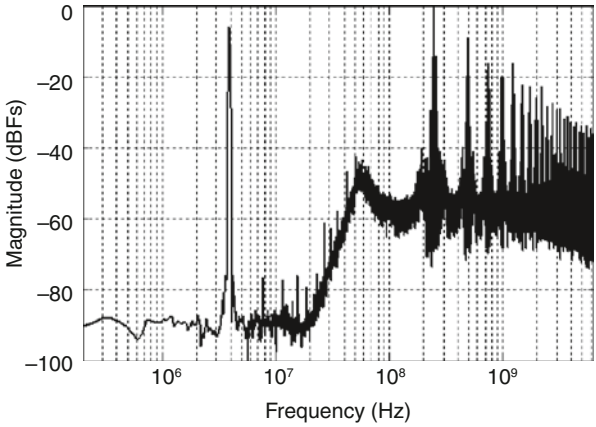


Fig. 7 Measured output spectrum of the modulator

Valuable insights can be gained through simulations using the MATLAB model of the ADC with additive Gaussian white phase noise in the clock waveform. Simulations with various values of RMS clock jitter up to 3.5 psecs show that that the SNR limit of the PWM modulator is at least 10 dB higher than with conventional feedback using an RZ pulse shape [13]. This is attributed to the low-sensitivity of the PWM pulse width to clock jitter.

2.6 Experimental Results

An ADC prototype was fabricated in TI 65 nm digital CMOS technology. A low-jitter clock for the ADC was generated with an on-board SAW oscillator. The data captured from the logic analyzer represents the quantized pulse timing edges from which the PWM waveform can be reconstructed in MATLAB. Figure 7 shows the measured modulator's output spectrum. The ADC's dynamic range is found to be 68 dB. The peak SNR and SNDR are approximately 62 and 60 dB, respectively. The peak total harmonic distortion (THD) is about 67 dB and occurs at an input level of −6 dBFs.

Table 2 compares the performance of the described ADC with that of state of the art ΔΣ ADCs. The conventional figure of merit $\text{FOM} = \text{Power}/(2 \cdot \text{BW} \cdot 2^{\text{ENOB}})$ is used as a comparison metric. Since the output data rate of the described ADC is much lower than that of other ADCs, it permits the use of a slower clock frequency for the digital decimation filters following the modulator.

Table 2 Performance comparison of the ADC with the state of the art

	[3]	[5]	[7]	This work
SNDR (dB)	70	69	55	60
Power (mW)	27.9	56	38	10.5
Area (mm^2)	1.0	0.5	0.19	0.15
Output rate (MSPS)	420	680	950	250
FOM (fJ/step)	270	298	2058	319

3 A 5th-Order Continuous-Time Lowpass ΔΣ Modulator [18]

In this section, an alternative design technique with PWM feedback is presented and an example ADC design is discussed which has 25 MHz bandwidth and over 67 dB SNDR in a vanilla 0.18 μm CMOS technology.

3.1 Multi-Bit ADC Architecture

Figure 8 shows the fully-differential 5th-order lowpass ΔΣ modulator (feedforward architecture) with 25 MHz signal bandwidth and a sampling frequency of 400 MHz. A local feedback composed of a standard 3-bit NRZ DAC is used to alleviate the effects of excess loop delay, but the device mismatch tolerance is much higher for this DAC in front of the quantizer because its nonlinearity error is noise-shaped. To achieve the noise-shaping over a wide frequency range (~DC to 25 MHz), a 5th-order quasi-linear phase inverse Chebyshev lowpass filter is employed; which consists of two cascaded 2nd-order lowpass sections and a lossy integrator. The respective 3-dB frequencies are set to 24.5, 16.7 and 5.71 MHz. The summing amplifier (Σ) couples all feedforward paths to the quantizer input. A 3-bit two-step quantizer is employed to minimize the processing delay in combination with the PWM feedback DAC.

Instead of employing the PWM in the signal path as the modulator in Sect. 2, the level-to-PWM converter in this alternative architecture is implemented in the feedback path to convert the digital codes from the 3-bit quantizer to time-domain PWM signals compatible with the single-bit current steering DAC. This realization also avoids performance degradation originating from unit current source mismatches linked to conventional multi-bit DACs. As shown in Fig. 8, the pulse shapes of feedback PWM signals are arranged as symmetric as possible within a clock period to minimize power of aliasing tones [17, 19]. A fully symmetric DAC requires increasing the number of clock phases; and the effects of the asymmetric pulses are analyzed in Sect. 3.5. To generate the clock's seven evenly distributed phases (Φ_1–Φ_7), a LC tank VCO and complementary injection-locked frequency divider (CILFD) were combined for low-jitter performance with low power consumption [20].

3.2 Design Considerations

3.2.1 Jitter Sensitivity

The drawback of employing multi-phase time-domain signals is increased sensitivity to jitter noise because of larger and more frequent DAC output transitions

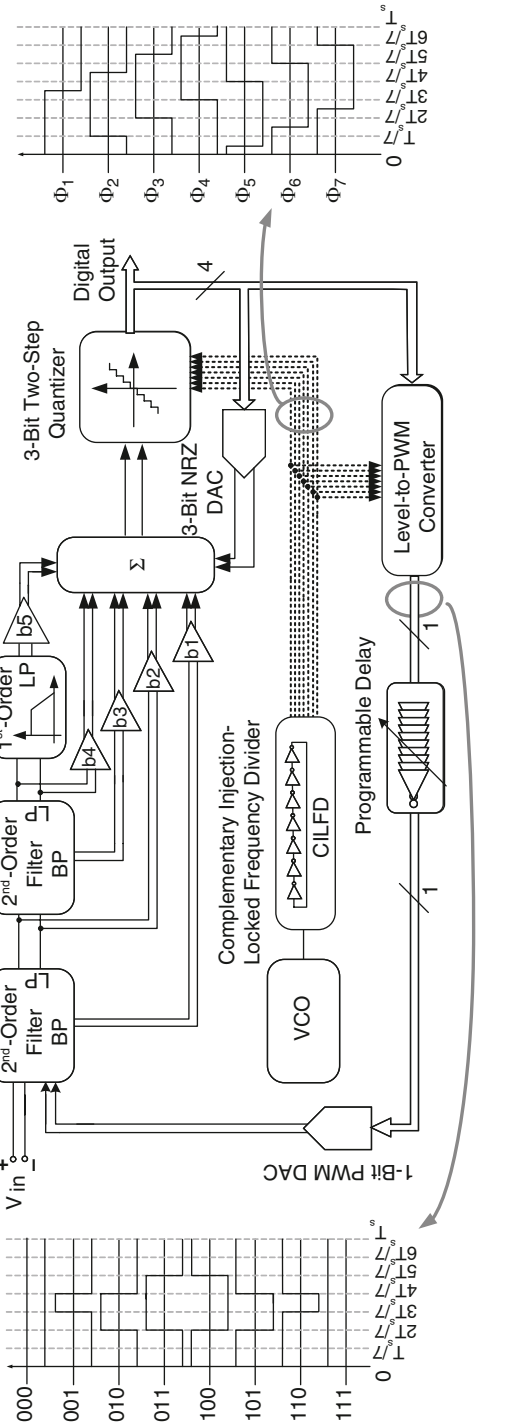


Fig. 8 System architecture of the 5th-order CT modulator

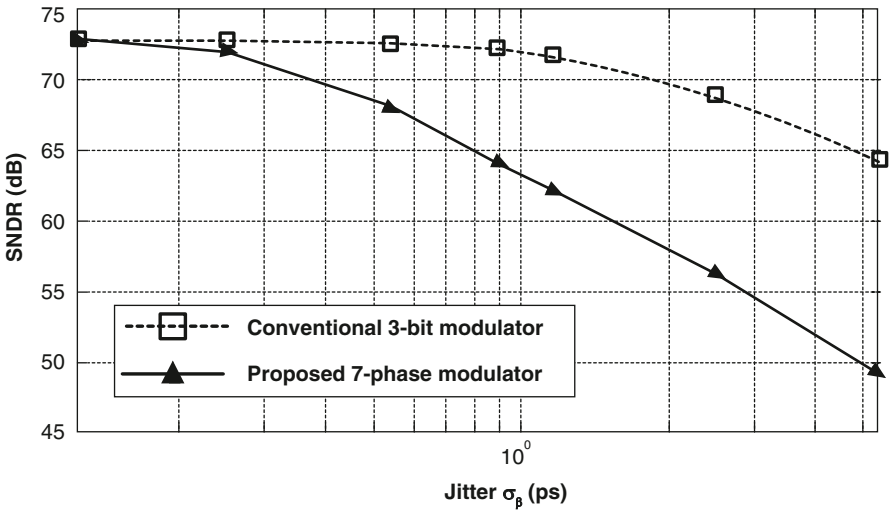
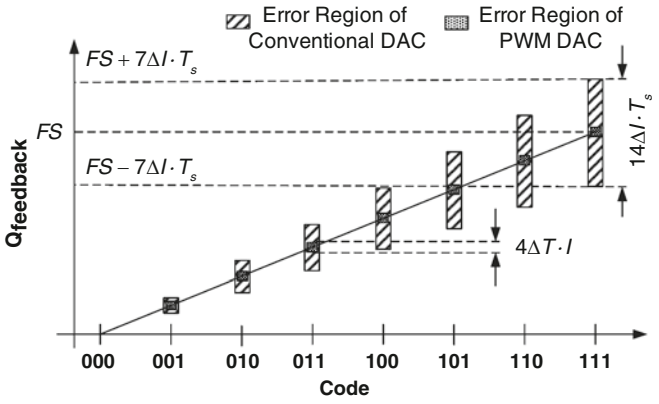


Fig. 9 SNR (SQNR+SJNR) vs. jitter (uncorrelated for all clock phase) of the modulator with 7-phase PWM DAC compared to the typical 3-bit DAC

compared to a conventional 3-bit non-return-to-zero (NRZ) DAC. Figure 9 shows the simulated SNR performance vs. clock jitter for the modulator under examination and a 3-bit modulator with conventional NRZ DAC (400 MHz sampling). If the clock jitter for the 7-phases is uncorrelated (worst-case assumption), then the integrating window is affected; and with $\sigma_\beta \approx 0.5$ ps, the SNR of the modulator with PWM DAC is 5 dB lower than that of a typical 3-bit NRZ DAC at 400 MHz. However, in practice, the seven phases are generated with the same master clock. Hence, they are highly correlated, making the architecture more robust to clock jitter. It can be shown that jitter effects are shaped by a function in the form of $(1-Z^{-N})$, where N (>1) is related to the quantizer code. The single-element PWM DAC is not affected by SNDR reduction from unit current source mismatches as the 3-bit multi-element NRZ DAC. From Fig. 9, the worst-case clock jitter requirement for SNDR > 68 dB with the described modulator is $\sigma_\beta < 0.54$ ps.

3.2.2 Static Device Mismatch

The nonlinearity of the PWM DAC due to static timing mismatches can be assessed from a feedback charge error comparison relative to the conventional 3-bit DAC. Figure 10 visualizes the worst-case peak-to-peak charge errors for each code, which are resultants of static mismatch ΔI_i for each current cell in the conventional DAC and static timing error ΔT_j of clock phase Φ_j in the PWM DAC. ΔT_j originates from the static CILFD mismatches and unequal propagation delays due to routing parasitics. Notice that the errors depend on mismatches in up to seven unit elements of the conventional DAC, but only up to two timing phases with



$\lvert\Delta Q\rvert$	000	001	010	011	100	101	110	111
Conventional DAC	0	$\Delta I_1 \cdot T_s$	$\sum_{i=1}^{2} \Delta I_i \cdot T_s$	$\sum_{i=1}^{3} \Delta I_i \cdot T_s$	$\sum_{i=1}^{4} \Delta I_i \cdot T_s$	$\sum_{i=1}^{5} \Delta I_i \cdot T_s$	$\sum_{i=1}^{6} \Delta I_i \cdot T_s$	$\sum_{i=1}^{7} \Delta I_i \cdot T_s$
PWM DAC	0	$(\Delta T_4+\Delta T_5) \cdot I$	$(\Delta T_3+\Delta T_5) \cdot I$	$(\Delta T_3+\Delta T_6) \cdot I$	$(\Delta T_3+\Delta T_6) \cdot I$	$(\Delta T_3+\Delta T_5) \cdot I$	$(\Delta T_4+\Delta T_5) \cdot I$	$2\Delta T_1 \cdot I$

Fig. 10 Relative 3-bit DAC linearity error comparison: conventional vs. PWM

the PWM scheme. Assuming equal mismatches ($\Delta I_i=\Delta I$, $\Delta T_j=\Delta T$) yields worst-case errors of $\pm 7\Delta I \cdot T_s$ and $\pm 2\Delta T \cdot I$ for conventional and PWM DACs, respectively. Letting $\delta_{\%I}=\Delta I/(I/7)$ and $\delta_{\%T}=\Delta T/(T_s/7)$ be the percent standard deviations of the mismatches in each case, the worst-case errors are $\Delta Q_{conv.\text{-}worst}=\pm 7\delta_{\%I} \cdot (I/7) \cdot T_s$ and $\Delta Q_{PWM\text{-}worst}=\pm 2\delta_{\%T} \cdot I \cdot (T_s/7)$. Monte Carlo post-layout simulations including delay mismatches in all clock phases showed that $\delta_{\%T}=0.16\%$ as a result of the synchronizing effect from the injection-locking. Since $\delta_{\%I}$ is typically 0.5% with good layout practices for a standard DAC, the anticipated worst-case linearity error of the PWM DAC is favorably lower.

Assuming that two timing mismatches are accumulated in the case of the PWM-based ADC, all mismatches in the conventional realization are accumulated, and errors are uncorrelated in both cases. Thus, the induced third harmonic distortion (HD3) comparison ratio can be derived, resulting in:

$$\frac{HD3_{PWM}}{HD3_{conventional}} \cong \left(\sqrt{\frac{2}{N}}\right)\left(\frac{\delta_{\%T}}{\delta_{\%I}}\right), \tag{7}$$

where N is the number of DAC levels. For N = 7 and the aforementioned distributions, the linearity of the proposed PWM DAC outperforms the conventional DAC by 15 dB based on (7).

3.3 Filter and Summing Amplifier

Active-RC filter topologies are used in the loop filter and the summing amplifier to exploit their linearity advantage. Since the noise and distortion introduced by the first biquad stage is most critical for overall performance, the first two-integrator-loop active-RC filter in Fig. 8 was designed with sufficient linearity as well as relatively small input resistors, large integrating capacitors, and amplifiers having gain greater than 40 dB at 25 MHz for adequate thermal noise levels. A fully-differential two-stage topology of the single-ended equivalent diagram shown in Fig. 11 was implemented, where feedforward compensation has been adopted [21]. The first stage of the amplifier was designed to have high gain and a dominant pole determined by the parasitic capacitance (C_{L1}) and overall output resistance (R_{L1}) of the first stage. The second and feedforward stages are optimized for large gain up to 25 MHz; design details can be found in [18]. Simulations showed that amplifier gain larger than 44 dB at 24.5 MHz and IM3 below −73.5 dB with 400 mV_{p-p} output swing are achieved in all process and temperature corners; better linearity figures can be obtained (if needed) by employing linearization techniques [22]. Capacitor banks with ±30% tuning range were utilized for compensation of the filter's time constant variations. This programmability would allow for automation of performance tuning such as the one proposed in [23].

The summing amplifier is a critical analog block of the architecture with a local feedback path around the quantizer. For precise equivalence between discrete-time and continuous-time loop transfer functions it is required to maintain exactly one sampling period delay in the local feedback path; usually this path defines the high frequency behavior of the loop and ensures loop stability. The summing amplifier

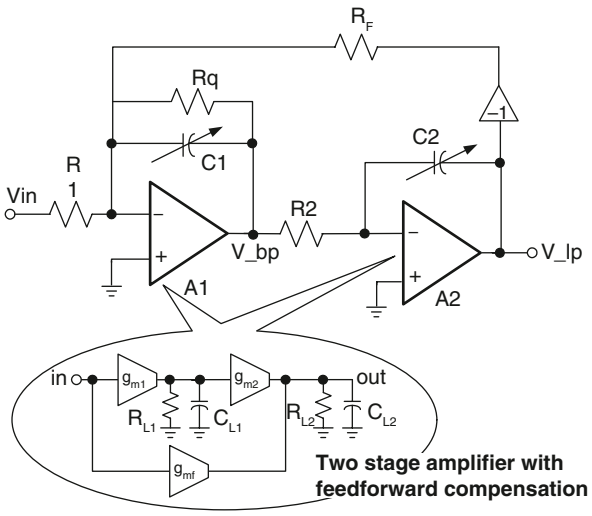


Fig. 11 Single-ended equivalent block diagram of the biquad and amplifier

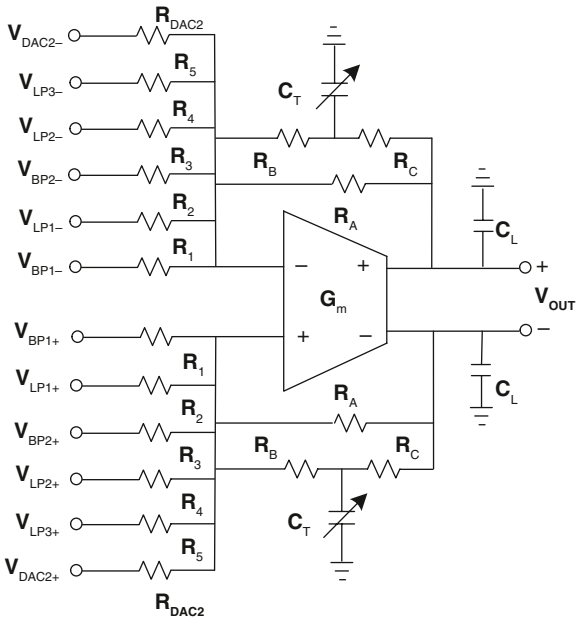


Fig. 12 Summing amplifier stage

displayed in Fig. 12 contains a bandwidth extension scheme. The feedback resistor is split into two pieces, and one of them is replaced by a T-RC network. This feedback network creates a zero-pole pair, which introduces positive excess phase in the overall transfer function to adjust for the group delay of the summing node. Capacitor C_T is tuned to optimize the loop delay in the local feedback path consisting of the summing amplifier, quantizer, and secondary NRZ DAC.

3.4 Three-Bit Two-Step Quantizer

In comparison to conventional flash architectures, multi-step quantization can reduce area and power consumption when multiple clock phases/cycles are available [24]. The presented architecture is a four-step quantizer that combines a 1-bit MSB decision with a second step during which the input is successively compared to reference levels.

As illustrated in Fig. 13a, the quantizer utilizes the seven clock signals to control four sequential comparison instants (τ_1–τ_4), which cuts the number of comparators from seven to four with respect to a typical 3-bit flash ADC. The four-step process makes the MSB available after the first step, creating timing margin for the digital control logic that sets up the PWM DAC. Successive approximations during the second step resolve the remaining bits that are processed by the level-to-PWM converter. Similarly to the combination of the PWM generator and TDC in Sect. 2, the 1-bit DAC is also driven by a PWM waveform. However, in the approach presented here, successive approximation techniques are employed. The successive algorithm

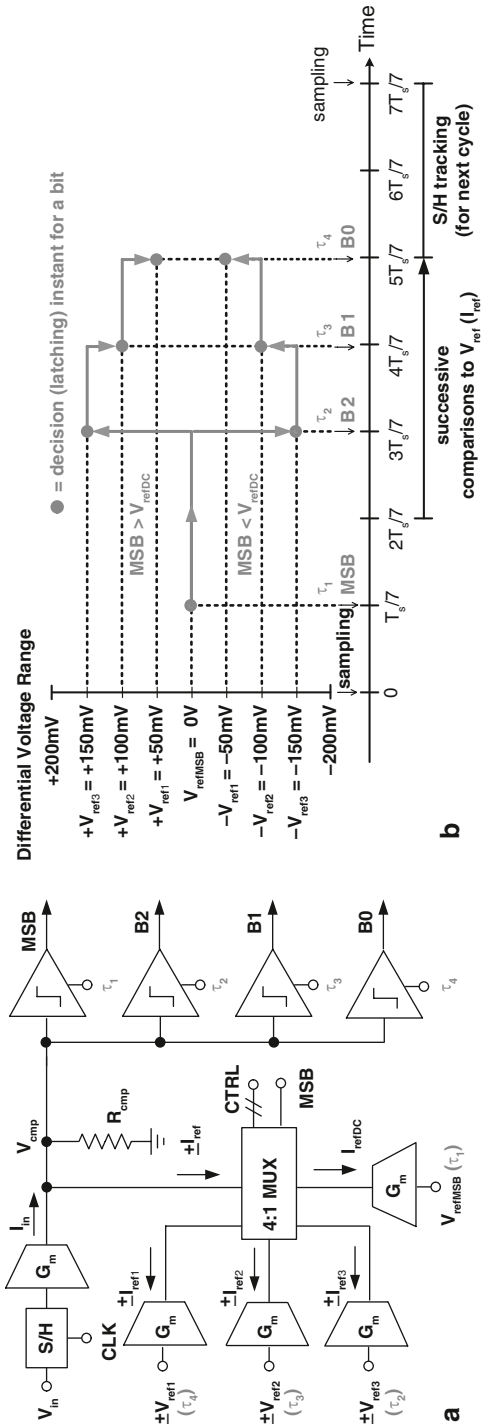


Fig. 13 Quantizer: **a** single-ended block diagram, **b** timing of the decisions

only has one MSB and three LSB quantization steps; the comparison to discrete reference levels is a simple alternative that also gives the option to calibrate each level individually if necessary. The quantizer and its corresponding timing diagram is shown in Fig. 13b and operates as follows. The differential input signal V_{in} is sampled by a S/H circuit at the beginning of the master clock having a period T_s, and then it is converted to current I_{in} via a G_m. First, the MSB is resolved after τ_1 seconds by comparing I_{in} to the current from V_{refMSB} applied to an identical G_m stage. Depending on the timing control bits (CTRL) and the MSB decision, a multiplexing configuration (MUX) is utilized to compare I_{in} to current I_{ref} derived from the appropriate differential reference voltage ($\pm V_{ref1} \ldots \pm V_{ref3}$) during each subsequent instant (τ_2–τ_4). The order of the subranging comparisons and output bits was chosen based on the timing needs in the multi-phase DAC control circuitry because larger signal magnitudes require DAC feedback pulse changes early in the next clock cycle. Comparison resistor (R_{cmp}) converts the difference in currents into a positive or negative voltage. Thus, the latch states represent the instant in time at which the discrete reference ramp has crossed the input signal level, providing the time-domain information for the PWM DAC.

3.5 *Level-to-PWM Converter*

Figure 14 presents the fully-differential block diagram of the level-to-PWM converter. The pulse waveforms are generated with SR latches. In every sampling period, the AND gates determine the appropriate pulse shape to be passed through the 5-input OR gate to the 1-bit DAC according to the output codes of the quantizer. A programmable delay block constructed with a series of digital inverters and a 3-bit MUX is included in the feedback path to avoid SNR degradation by ensuring that the excess loop delay is within 5%. The delayed differential PWM signals are syn-

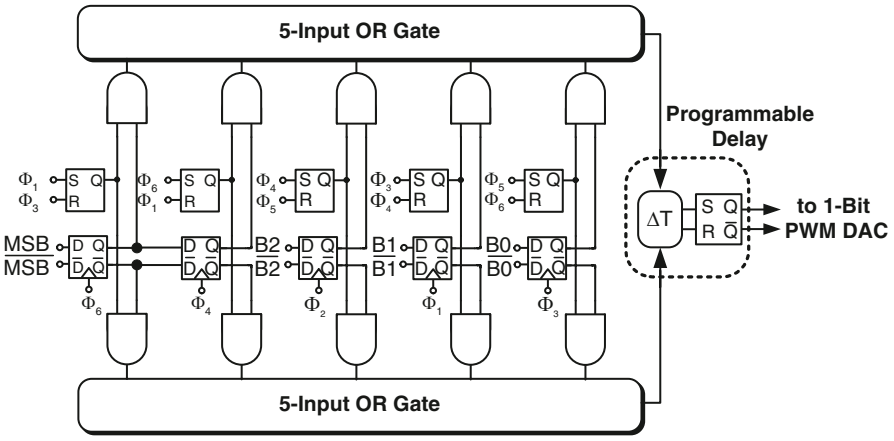


Fig. 14 Implementation of the level-to-PWM converter

chronized by an additional SR latch that precedes the 1-level PWM DAC. The layout as well as routing were planned carefully to minimize the timing mismatches.

The mapping from a conventional multi-level digital signal to a PMW signal generates harmonic distortions, which may degrade the modulator performance. Figure 15 shows the output spectra of behavioral simulations with symmetric and asymmetric PWM pulse shapes. The pseudo-symmetric pulse shape (generated with 7 phases) results in 2 dB SNDR degradation because more out-of-band noise is aliased back by the nonlinear effects. The out-of-band distortions with the pseudo-symmetric pulse shape are 5 dB larger than with the perfectly symmetric pulse shape, which can be observed in the zoomed-in plot.

These results demonstrate that the modulator with triangular-wave PWM pulse shape has good in-band linearity performance, even with the pseudo-symmetric pulse shape.

If a sinusoidal input signal is used, the quantization distortion for an open-loop 3-bit ADC is in the range of −27 dB. Since the quantizer is in the feedforward path

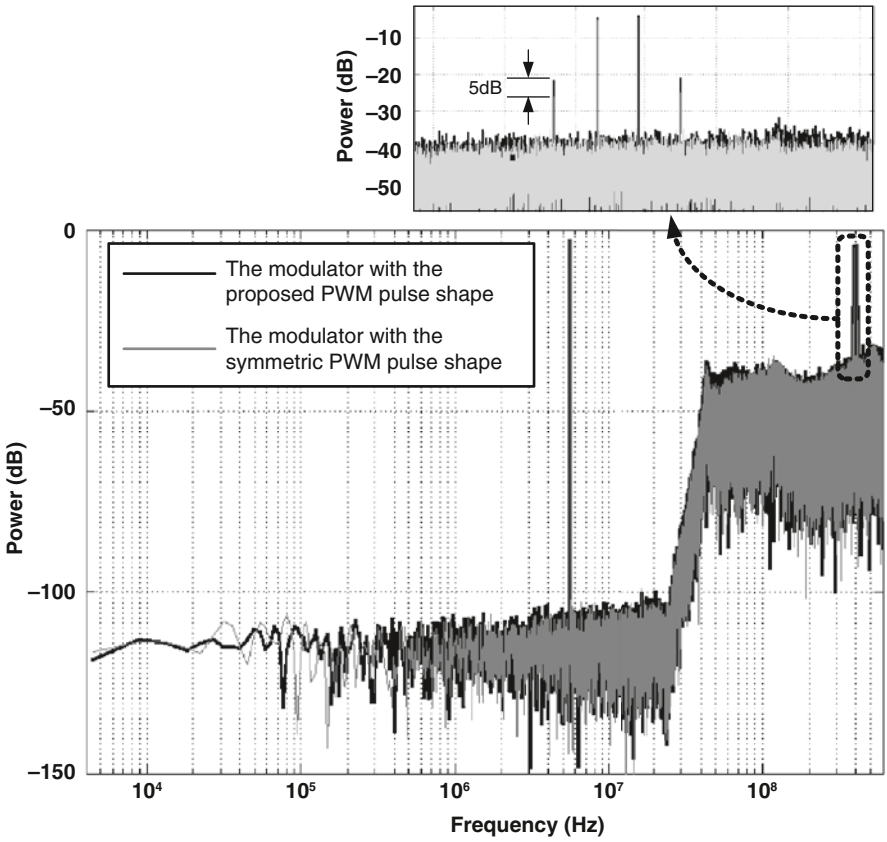


Fig. 15 Spectrum comparison: Asymmetric pulse shape vs. symmetric pulse shape

of the ADC, this distortion is attenuated by a factor $(1+\text{Loop\_Gain})^3$, leading to HD3 < −80 dB. DAC non-idealities affect system performance in a different way. Since the DAC is in the feedback path, its noise and nonlinearity directly affect the modulator's performance. The mapping from a conventional multi-level signal to a PWM signal generates harmonic components at multiples of the fundamental signal, harmonics related to the sampling clock (PWM pulse fundamental), and products due to the intermodulations between the harmonics. The power of these components can be obtained from equation (5). It is interesting to compute the HD3 for the sinusoidal-input-to-PWM-output conversion (symmetric and asymmetric), which is approximately given by [17],

$$HD3 \cong \left(\frac{0.88M^2}{4}\right)\left(\frac{f_{signal}}{f_{clock}}\right)^2, \tag{8}$$

where M is the modulation index. For M = 0.5 (−6 dBFS) and f_{signal}/f_{clock} = 20 MHz/400 Mz, the HD3 is predicted to be around −77 dB.

3.6 Clock Generator

The clock generation is performed with a 2.8 GHz VCO whose differential output signal is injected into a divide-by-7 CILFD, providing 7-phase outputs at 400 MHz. As shown in Fig. 16, the divide-by-7 CILFD is implemented with seven ring-type inverter stages composed of an upper tail-injection transistor, M_{pt}, a bottom tail-injection transistor, $M_{nt,}$ and inverter transistors (M_p, M_n).

The phase noise of the ring oscillator outputs is mainly determined by the lower phase noise of the injected VCO signal when the stages are locked [20]. With *n* inverters in a ring, the phase noise of a CILFD can be approximated as

$$PN_{(CILFD)} \cong PN_{(VCO)} - 10\,log_{10}(2n+1)^2. \tag{9}$$

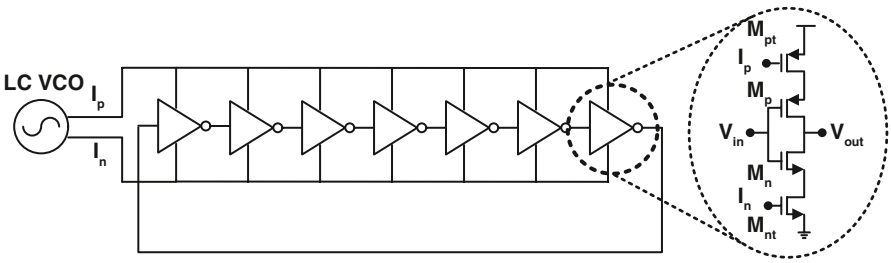


Fig. 16 Complementary injection-locked frequency divider (CILFD) diagram

The injection signal from the 2.8 GHz VCO presents phase noise of −119 dBc/Hz at 1 MHz offset frequency, and the divide-by-7 CILFD outputs show phase noise of −136 dBc/Hz at 1 MHz offset.

3.7 Simulated and Experimental Results

The architecture was modeled in SIMULINK and has a simulated SQNR of 73.4 dB in 25 MHz bandwidth, as shown in Fig. 17. It was fabricated in Jazz Semiconductor 0.18 μm 1P6M CMOS technology. Excluding the VCO, the power budget from a 1.8 V supply is 44 mW for the modulator core, 2.5 mW for the CILFD, and 1.5 mW due to clock buffers. Of this power, 27 mW (56%) is consumed by the quantizer and level-to-PWM converter.

Figure 18 plots the measured SNR and SNDR for different input signal powers obtained by injecting the single input signal at 5 MHz. The peak SNDR is 67.7 dB when signal power is −2.2 dBFS. The third-order harmonic distortion (HD3) in this case is 78 dB below the test tone, which demonstrates the high linearity properties of both loop filter and DAC. The linearity performance was further characterized by injecting two tones with 2 MHz separation, each having a power of −5 dBFS. The IM3 from two-tone tests at different frequency locations is better than 72 dB from ~DC up to 25 MHz.

The blocker rejection capability of the modulator was also measured by employing a −10 dBFS input power at 390 MHz (10 MHz offset from the clock frequency) to emulate the appearance of a blocker from another channel. The power level of this interference measured at 10 MHz is attenuated down to −66 dBFS, achieving 56 dB blocker rejection. Table 3 shows a comparison between the described modulator archi-

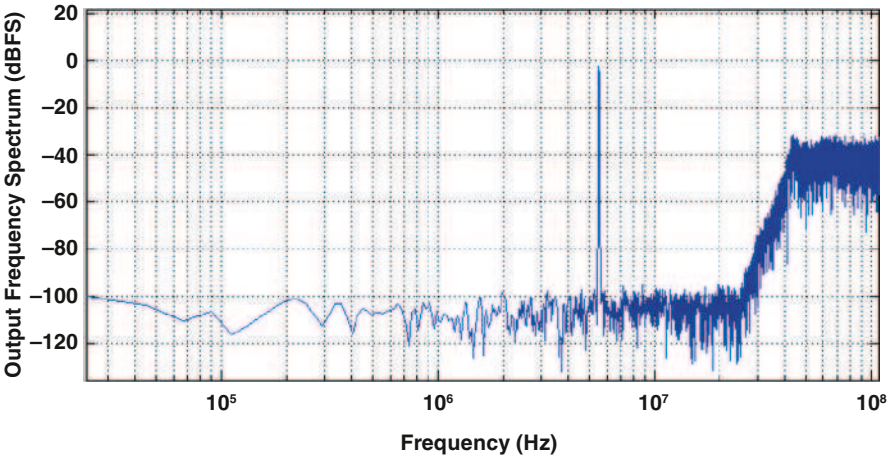


Fig. 17 Simulated output spectrum of the proposed modulator

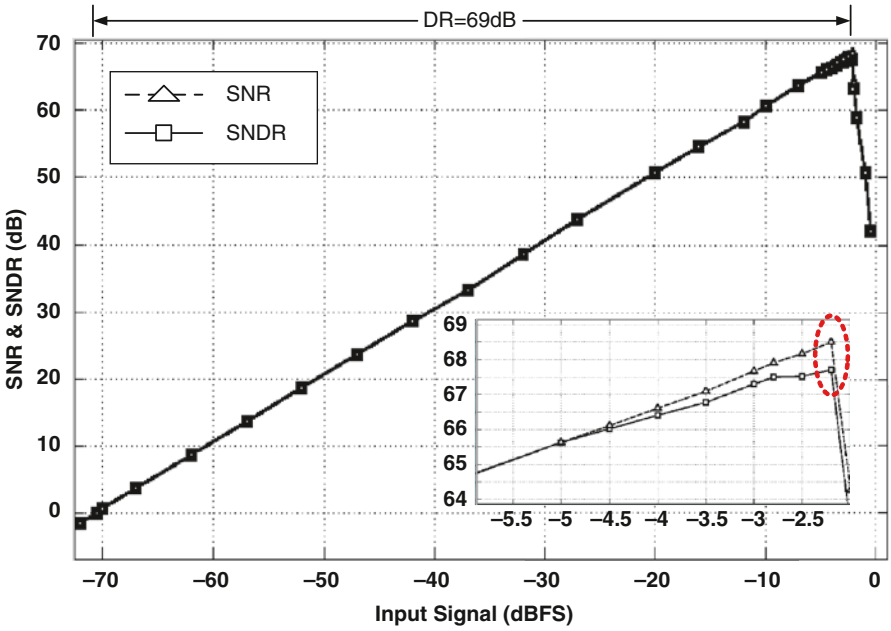


Fig. 18 Measured SNR and SNDR vs. input signal power

tecture and recently reported modulators based on the FOM = Power/($2 \cdot BW \cdot 2^{ENOB}$). Although fabricated in an economical technology, the achieved 444 fJ/bit FOM of the discussed modulator core is competitive with the current state of the art while providing high rejection to strong blockers. A FOM improvement is anticipated if the solution is exported to deep submicron technologies, which would lower the quantizer and level-to-PWM converter power as a result of more efficient digital circuitry.

Table 3 Comparison with previously reported LP ΔΣ ADCs

Reference	Technology (nm CMOS)	f_{clock} (MHz)	BW (MHz)	Order	Peak SNDR (dB)	Power (mW)	FOM (fJ/bit)
[3] ISSCC 2008	90	420	20	4	70	28[b]	271[d]
[5] ISSCC 2007	90	340	20	4	69	56[c]	608
[6] JSSC 2006	130	640	20	3	74	20[b]	122
[8] JSSC 2008	130	950	10	2	72	40[a]	500
[12] ISSCC 2009	65	250	20	3	60	10.5[b]	319
[25] ISSCC 2008	180	640	10	5	82	100[b]	487
This work	180	400	25	5	67.7	44[b]	444

[a] Includes clock generation circuitry
[b] For modulator circuitry only
[c] Includes digital calibration of RC spread & noise cancellation filter
[d] Discrete-time modulator (would require anti-aliasing filter for comparable blocker rejection)

4 Conclusion

Various design issues for next generation wideband continuous-time ΔΣ modulators have been addressed in this overview chapter through the discussion of two example designs. Both of these architectures are utilizing time-based quantization and feedback techniques, which were developed with the aim to adapt to technology scaling. Foremost, the schemes provide multi-bit feedback with single-element DACs, circumventing the matching problems of typical multi-element DACs that become worse as scaling continues. In parallel, time-based methods take advantage of increased clock frequencies and improved timing resolution. These type of ADCs present viable alternatives for high-performance applications when low-jitter clock signals can be generated.

Acknowledgements The authors would like to recognize that these ADCs were developed with the collaboration of Prof. E. Sanchez-Sinencio, M. Elsayed, E. Pankratz, Y. C. Lo, V. Gadde, and V. Periasamy. The authors appreciate the chip fabrications sponsored by Texas Instruments and Jazz Semiconductor.

References

1. J.A. Cherry, W.M. Snelgrove, *Continuous-Time Delta-Sigma Modulators for High-Speed A/D Conversion: Theory, Practice and Fundamental Performance Limits*. (Kluwer, New York, 2002)
2. S.R. Norsworthy, R. Schreier, G.C. Temes, *Delta-Sigma Data Converters: Theory, Design, and Simulation*. (Wiley-IEEE Press, New York, 1996)
3. P. Malla, et al., A 28 mW spectrum-sensing reconfigurable 20 MHz 72dB-SNR 70dB-SNDR DT ΔΣ ADC for 802.11n/WiMAX receivers, in *IEEE ISSCC Digest of Technical Papers*, San Francisco, CA, Feb 2008, pp. 496, 497
4. G. Mitteregger, et al., A 14 b 20 mW 640 MHz CMOS CT ΔΣ ADC with 20 MHz signal bandwidth and 12 b ENOB, in *IEEE ISSCC Digest of Technical Papers*, San Francisco, CA, Feb 2006, pp. 131–140
5. L.J. Breems, et al., A 56 mW CT quadrature cascaded ΣΔ modulator with 77 dB DR in a near zero-IF 20 MHz band, in *IEEE ISSCC Digest of Technical Papers*, San Francisco, CA, Feb 2007, pp. 238, 239
6. G. Mitteregger, et al., A 20-mW 640-MHz CMOS continuous-time ΣΔ ADC with 20-MHz signal bandwidth, 80-dB dynamic range and 12-bit ENOB. IEEE J. Solid-State Circuits **41**(12), 2641–2649 (2006)
7. M.Z. Straayer, M.H. Perrott, A 10-bit 20 MHz 38 mW 950 MHz CT ΣΔ ADC with a 5-bit noise-shaping VCO-based quantizer and DEM circuit in 0.13 u CMOS, in *Proceedings of IEEE VLSI Circuits Symp*., Kyoto, Japan, June 2007, pp. 246, 247
8. M.Z. Straayer, M.H. Perrott, A 12-bit, 10-MHz bandwidth, continuous-time ΣΔ ADC with a 5-bit, 950-MS/s VCO-based quantizer. IEEE J. Solid-State Circuits. **43**(4), 805–814 (2008)
9. R.E. Radke, A. Eshraghi, T.S. Fiez, A 14-bit current-mode ΣΔ DAC based upon rotated data weighted averaging. IEEE J. Solid-State Circuits. **35**, 1074–1084 (2000)
10. E. Fogleman, I. Galton, A dynamic element matching technique for reduced-distortion multibit quantization in delta-sigma ADCs. IEEE Trans. Circ. Syst. II. **48**(2), 158–170 (2001)
11. Y. Tsividis, Digital signal processing in continuous time: A possibility for avoiding aliasing and reducing quantization error, *in Proceedings of International Conference on Acoustics, Speech, Signal Processing*, Montreal, Canada, May 2004, vol. II, pp. 589–592

12. V. Dhanasekaran, M. Gambhir, M. Elsayed, E. Sanchez-Sinencio, J. Silva-Martinez, C. Mishra, L. Chen, and E. Pankratz, A 20 MHz BW 68 dB DR CT ΔΣ ADC based on a multi-bit time-domain quantizer and feedback element, in *IEEE ISSCC Digest of Technical Papers*, Feb 2009, pp. 174, 175
13. V. Dhanasekaran, Baseband analog circuits in deep-submicron CMOS technologies targeted for mobile multimedia, Ph.D. Dissertation, Texas A&M University, College Station, TX, Aug 2008
14. H. Frank, U. Langmann, Excess loop delay effects in continuous-time quadrature bandpass sigma-delta modulators, in *Proceedings of IEEE International Symposium on Circuits and Systems* (*ISCAS*), vol. 1, May 2003, pp. 1029–1032
15. Y. Arai, T. Baba, A CMOS time to digital converter VLSI for high-energy physics, in *IEEE VLSI Circuits Digest of Technical Papers*, Tokyo, Japan, June 1988, pp. 121, 122
16. S. Verma, et al., A unified model for injection-locked frequency dividers, IEEE J. Solid-State Circuits. **38**(6), 1015–1027 (2003)
17. D.G. Holmes, T.A. Lipo, *Pulse width modulation for power converters: Principles and practice.* (IEEE Press, Piscataway, NJ, 2003)
18. C.-Y. Lu, M. Onabajo, V. Gadde, Y.-C. Lo, H.-P. Chen, V. Periasamy, J. Silva-Martinez, A 25 MHz bandwidth 5th-order continuous-time lowpass sigma-delta modulator with 67.7 dB SNDR using time-domain quantization and feedback. IEEE J. Solid-State Circuits **45**(9), 1795–1808 (2010)
19. F. Colodro, A. Torralba, New continuous-time multibit sigma-delta modulators with low sensitivity to clock jitter. IEEE Trans. Circ. Syst. I. **56**(1), 74–83 (2009)
20. Y.-C. Lo, H.-P. Chen, J. Silva-Martinez, S. Hoyos, A 1.8 V, sub-mW, over 100% locking range, divide-by-3 and 7 complementary-injection-locked 4 GHz frequency divider, in *Proceeding of IEEE Custom Integrated Circuits Conference* (*CICC*), Sept 2009, pp. 259–262
21. B.K. Thandri, J. Silva-Martinez, A robust feedforward compensation scheme for multi-stage operational transconductance amplifiers with no miller capacitors. IEEE J. Solid-State Circuits **38**, 237–243 (2003)
22. C.-Y. Lu, F. Silva-Rivas, P. Kode, J. Silva-Martinez, S. Hoyos, A 6th-order 200 MHz IF bandpass sigma-delta modulator with over 68 dB SNDR in 10 MHz bandwidth. IEEE J. Solid-State Circuits. (to be published (2010))
23. F. Silva-Rivas, C.-Y. Lu, P. Kode, B.K. Thandri, J. Silva-Martinez, Digital based calibration technique for continuous-time bandpass sigma-delta analog-to-digital converters. Analog Integr. Circ. S. **59**, 91–95 (2009)
24. B. Verbruggen, et al., A 2.2 mW 1.75 GS/s 5 bit folding flash ADC in 90 nm digital CMOS. IEEE J. Solid-State Circuits. **44**(3), 874–882 (2009)
25. W. Yang, et al., A 100 mW 10 MHz-BW CT ΔΣ modulator with 87 dB DR and 91dBc IMD, in *IEEE ISSCC Digest of Technical Papers*, Feb 2008, pp. 498–631

Oversampled DACs

Andrea Baschirotto, Vittorio Colonna and Gabriele Gandolfi

1 Introduction

High-resolution (≥16 bit) Digital-to-Analog Converters (DACs) have been the target of several important developments by using oversampled structures. In the first cases, oversampled DACs have been developed for the audio application, while recently the same concepts have been developed exploiting scaled technologies to process larger signal bandwidth with higher sampling frequency.

The overall architecture of an audio channel is shown in Fig. 1.

Digital audio signal is a stream of 16b-word at 44.1 kHz data-rate, whose frequency band is 20 Hz–20 kHz. The relative audio signal spectrum is shown in Fig. 2.

The Digital-to-Analog conversion of this signal presents two main problems. First, the filtering of the audio signal image: the image signal is starting at 24.1 kHz, i.e. 4.1 kHz away from the audio signal. A very sharp and very accurate filter frequency response is then needed. Second, the DAC accuracy in the order of 16 bit is beyond the achievable accuracy of DACs based on unit element arrays, which are then not sufficient for audio channels. These two problems are solved by means of oversampled ΣΔ DACs, whose generic architecture is shown in Fig. 3. The functionality of a ΣΔ DAC can be briefly described as follows. An interpolator increases the sampling frequency from F_s (=44.1 kHz) to $k \cdot F_s$, while all the images are cancelled, apart those at $k \cdot F_s$. This means that the frequency response accuracy requirements of the image rejection filter are strongly relaxed, as shown in Fig. 4.

Moreover, the DAC accuracy is solved by means of a digital ΣΔ modulator. The DAC interface is required to guarantee linearity performance in excess of the input signal quality, i.e. 16b for audio. A suitable solution is using a 1-bit DAC, which guarantees infinite linearity (for any two DAC output levels there is one line with infinite linearity). However, using a 1b DAC interface requires the 16b-word to be transformed into a 1b-word. This truncation operation introduces the truncation/quantization error, which could reduce signal accuracy. This is avoided by using a

A. Baschirotto (✉)
Department of Physics "G. Occhialini", University of Milano Bicocca, Milano, Italy
e-mail: andrea.baschirotto@unimib.it

H. Casier et al. (eds.), *Analog Circuit Design,*
DOI 10.1007/978-94-007-0391-9_12, © Springer Science+Business Media B.V. 2011

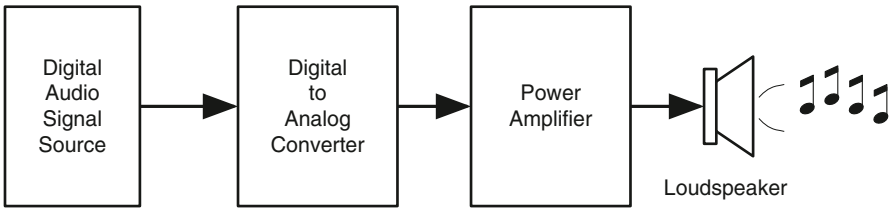
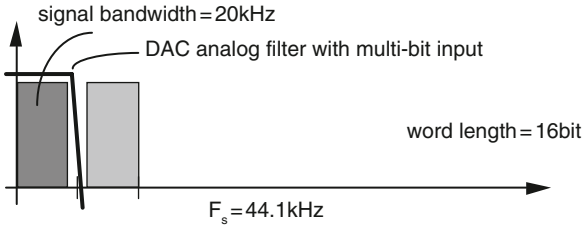


Fig. 1 Typical audio channel

Fig. 2 Audio signal spectrum



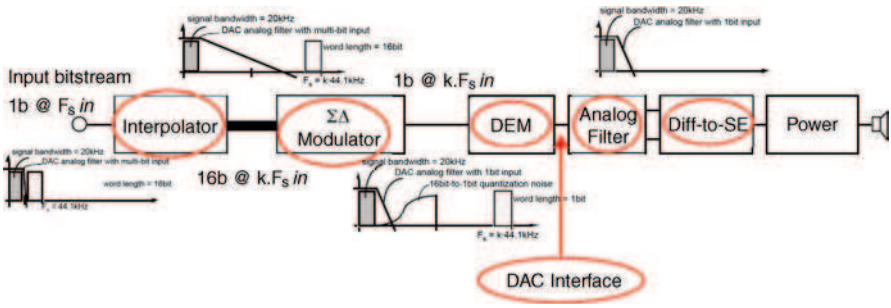


Fig. 3 Oversampled DACs architecture

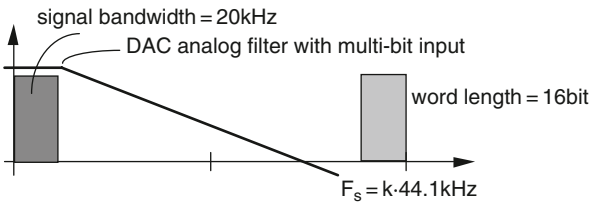


Fig. 4 Signal spectrum at the interpolator output

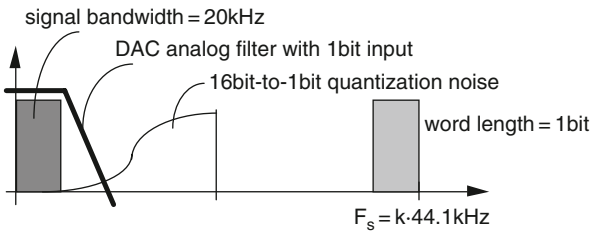


Fig. 5 Signal spectrum at the ΣΔ modulator output

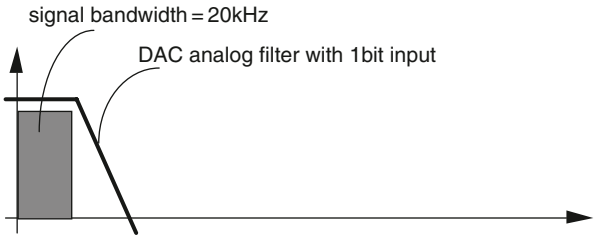


Fig. 6 Signal spectrum at the analog filter output

digital ΣΔ modulator, which shapes the truncation/quantization error spectrum in order to reduce in-band energy at the negligible cost of increasing out-of-band energy, exactly as an analog ΣΔ modulator shapes the quantization error. The resulting output spectrum is shown in Fig. 5.

The ΣΔ modulator bit-stream is then processed by the 1b-DAC and by the analog filter that rejects signal image and out-of-band truncation/quantization error, and produces an analog output signal, whose spectrum is shown in Fig. 6.

In the structure of Fig. 3 other functional blocks appear, like DEM, Diff-to-SE converter and Power stage. They will be eventually introduced later on.

2 Specification Analysis

Audio DACs are required to satisfy target specification relative to audio applications. Some of the most significant are here listed, in order to understand the audio DACs design choices.

2.1 In-band Dynamic Range (DR)

The in-band Dynamic Range (or generally Dynamic Range) is defined as the ratio between the signal amplitude and the Noise+Distortion (N+D) present in the audio band [20 Hz–20 kHz]:

$$DR = \frac{S_{MAX}}{N + D_{[20Hz-20kHz]}} > 96\,dB$$

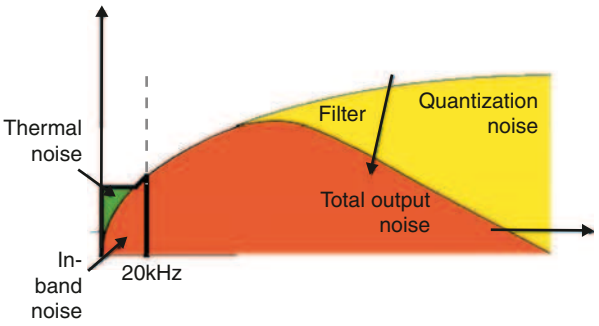


Fig. 7 Signal spectrum at the analog filter output

The in-band (N+D) is the combination of the truncation/quantization noise coming from the digital part and the electronic (thermal, 1/f, etc.) N+D coming from the analog filter. This second analog (N+D) contribution typically dominates and it has to be accurately minimized in the analog filter design (Fig. 7).

The main noise sources in an oversampled DAC are: the quantization noise, the 1/f noise of the active devices (opamp, and reference voltages), the thermal noise (opamp, reference voltages, and switches), the D/A interface noise (Jitter-correlated, ISI), and the folded noise (due to the distortion). The main distortion sources are: the analog filter non-linearity (opamp finite gain distortion, slew-rate, etc.), the SC-related distortion sources (switches, etc.), the eventual non-linearity in the output stages (diff-to-single and power stage), the Inter-Symbol-Interference (ISI) and Asymmetry in the DAC interface. The typical challenging requirement of THD < −80 dB (or more) is satisfied by using fully-differential structures. In the specification of an audio DAC, a THD at 0 dB_{FS} signal and at −60 dB_{FS} is required. The 0 dB_{FS} signal performance is demonstrating the signal processing quality for a large amplitude signal. On the other hand the −60 dB_{FS} test is due to the particular spectrum of the input signal coming from the preceding ΣΔM, as shown in Fig. 8. For a small input signal (−60 dBFS for the test) at f_{in}, there are several large spurious

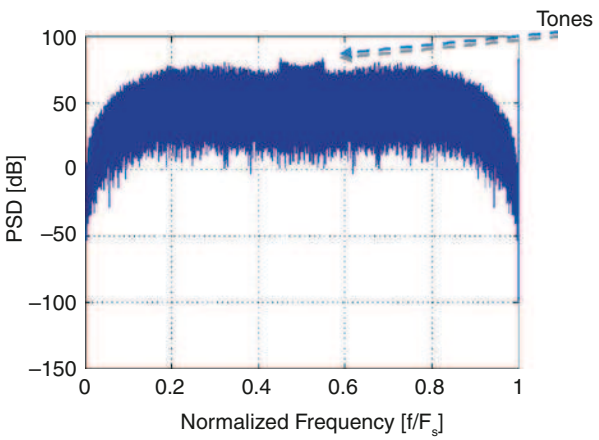


Fig. 8 Signal spectrum at the ΣΔ modulator output

tones around $F_s/2$, spaced by f_{in}. Any analog filter non-linearity produces intermodulation of these tones, which results in tones around dc spaced by f_{in}. This means that some tones at dc are present even in presence of a small amplitude input signal. Due to both these tests to be passed high-linearity DAC interfaces and analog filter are required in a high-performance DACs.

In the DR definition no specification is given for out-of-band noise: thus this specification leaves large freedom for the designer for the out-of-band performance. Since any additional filtering block increases in-band noise, the best solution for the DR performance would be a no-filter structure, that would results is large out-of-band noise [1, 2].

2.2 Low Power Consumption

Audio DACs are typically embedded in audio-codec, which are very complex SoC devices. Several ADC's and DAC's are typically present, together with some DSP for audio signal processing. A large device is then produced (an example is given in Fig. 9) that has to face power consumption and thermal dissipation issues. This means to force a strong power reduction for each block. This, of course, is also the case when a stand-alone DAC is embedded in a portable device, the power consumption has to be further minimized.

2.3 Single-Ended Output

Audio high-linearity performances are reached with fully-differential structures. However audio codecs require a single-ended output signal in order to drive grounded loads and to halve the number of external pins. This means that high-performance Differential-to-Single Ended converters are needed, in order to maintain also at the output single-ended node the same high-quality performance achieved with fully-differential structures.

2.4 Fully Integrability

The complete DAC channel would have to be fully integrated to avoid external components and/or additional filters.

2.5 Performance Robustness

DAC performance has to be guaranteed also in presence of some disturbs on the digital signal, like clock jitter typically present in commercially available audio ICs (a typical clock jitter value is 200 ps).

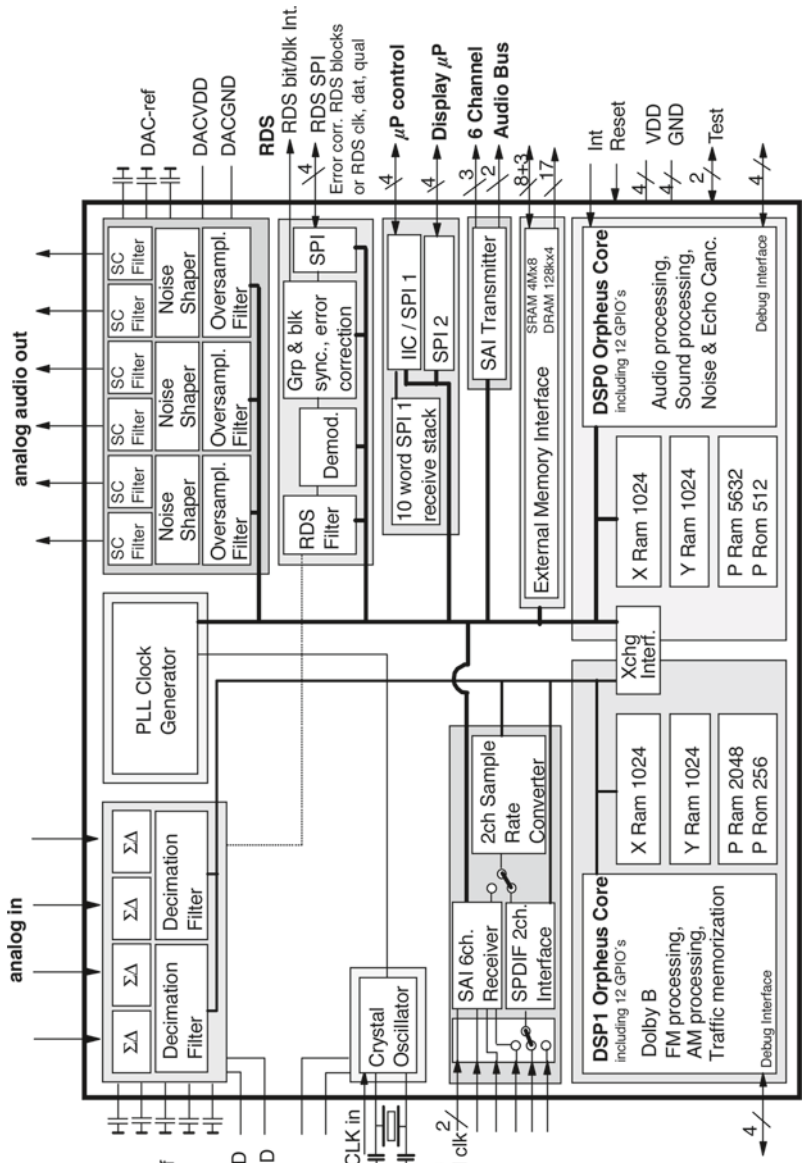


Fig. 9 Typical audio CoDec architecture

2.6 Accurate In-band and Out-of-band Frequency Response

Audio signal processing requires very accurate frequency response in the signal band. Maximum amplitude ripple is in the order of few tenth of dB in the [20 Hz–20 kHz] signal band. This results in two possible solutions: (1) design place filter poles far from the signal bandwidth to reduce their attenuation effect in the signal band. This is the case of continuous-time filters whose frequency response is typically very inaccurate; (2) use accurate analog filter frequency response implementation and compensate its drop in the signal band in the digital section (as shown in Fig. 10). This is the case of SC reconstruction filter implementation.

2.7 Out-of-band SNR_{out}

An additional parameter for qualifying an audio DAC is the out-of-band signal-to-noise ratio SNR_{out}, defined as the ratio of the full-scale signal and the total out-of-band noise.

$$SNR_{out} = \frac{S_{MAX}}{N_{\left[\frac{F_{Sout}}{2};2\cdot F_{Sout}\right]}}$$

The SNR_{out} is the result of the amount of out-of-band filtering implemented in the reconstruction filter. A small value of SNR_{out} corresponds to a significant amount of high-frequency energy, which may be dangerous for the coupling with the neighbor blocks and for the following output drivers. The SNR_{out} value can then be improved increasing the filter complexity at the cost of additional area and power (Fig. 11).

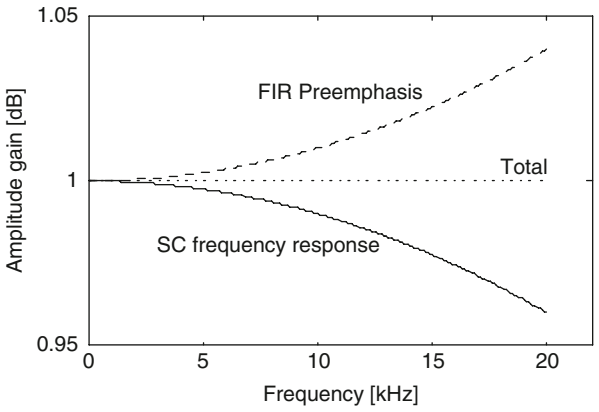


Fig. 10 Analog & Digital frequency response compensation

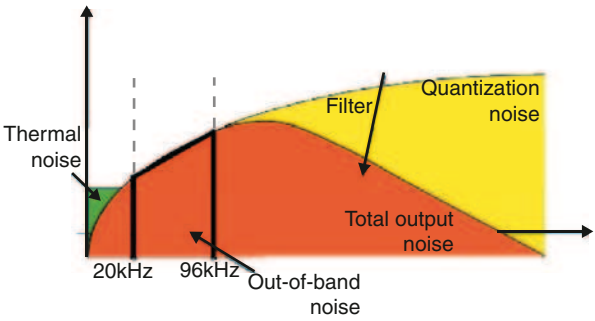


Fig. 11 Signal spectrum at the analog filter output

2.8 Output Power Stage

The output analog signal is typically fed to a power stage, which drives the external load (loudspeakers, headphones, etc.). In some recently developed audio DACs, a power stage is embedded in the analog filter and then the filter is able also to drive external load. Examples of these solutions are given in the following.

3 Audio-DAC Design Trade-off

In the design of an oversampled DAC the best trade-off between the requirements of the digital part (mainly the ΣΔM) and the analog part has to be found.

A starting point is the choice of the ΣΔM OSR. Since the audio signal data-rate is fixed (44.1 kHz), the OSR fixes both the digital data-rate and the analog filter operation frequency ($F_s = 44.1\ \text{kHz} \cdot \text{OSR}$).

In the first audio DACs, a single bit quantizer was used for linearity reason.

The target DR could be achieved with high-order (that is however avoided for stability reason) or high-OSR ΣΔMs. High-OSR ΣΔM has been the preferred solution. In this cases, a single-bit quantizer is used which requires challenging performance to the analog filter, as follows. Single bit quantizer presents large out-of-band noise (which requires higher filter order) and large input step (which requires high opamp slew-rate). Moreover, large OSR results in high data-rate (F_s) and then results in high-speed opamp in the analog filter. All these motivations lead to a large analog filter power. This solution was however very popular since it avoids any DEM algorithm that in non-scaled technology could be power and area hungry. Typical values for the OSR were 128 or 256.

Exactly for these reasons, as soon as scaled technologies allow efficient implementation of digital parts, a completely different approach has been preferred. Using a smaller OSR gives lower F_s and then the analog filter has to guarantee lower speed performance (opamp bandwidth, etc.) and consumes lower power. This however requires a high-order ΣΔM (this is avoided for stability reasons) or a multibit quantizer. A multibit quantizer has two positive effects on the analog filter: a lower

opamp slew-rate requirement and a lower out-of-band noise power. Thus the power consumption can be further decreased. However the additional cost of a multibit quantizer is the need of a Dynamic Element Matching (DEM) block to linearize the DAC interface, which could cost in term of circuit complexity, power consumption, and digital noise to the analog section. This costs in scaled technology is becoming negligible and then an increasing popular choice is using a low OSR value (typically 64, because a lower OSR would result in large quantization noise in the signal band) and a multibit quantizer. Several DEM algorithms have been developed. They have to optimize the trade-off between circuit complexity and performance. The most popular solution is the Data Weighted Average (DWA) algorithm and its modified versions. An important observation is that, since the DEM circuit is not embedded in a loop (as it is in A/D ΣΔMs) its latency time is not critical for structure stability. This gives some flexibility in the DEM algorithm implementation.

4 Digital-to-Analog Interface

The choice of the D-to-A interface regards the interface number of bits and the nature of the produced analog signal (current for a switched-current or switched-resistor DAC or charge for a switched-capacitor DAC).

4.1 Single-Bit vs. Multibit

The use of a single-bit SDM output data-rate features some key advantages: infinite linearity DAC (non need of DEM algorithms), and circuit simplicity. However there are some critical points like the presence of tones in the digital spectrum (single-bit ΣΔMs suffer from idle tones), the ΣΔM stability to be properly guaranteed. In addition, a large input step is applied to the analog filter and this requires large opamp slew-rate. Finally this exhibits high jitter sensitivity for current-mode D/A interface.

On the other hand, the use of a multibit quantizer guarantees a good SDM stability, no idle tones, small input step for small input signal and low jitter sensitivity. However a technique for linearizing the multi-bit DAC interface is needed. In literature there are two main class of linearizing technique: the calibration and the Dynamic Element Matching (DEM). The DEM is the most popular and is practically adopted in any multibit DAC. It consists in scrambling the choice of the unit elements of the DAC in order to average and null the non-linearity error. Several algorithms for the unit elements choice have been proposed. They have to be evaluated in consideration of the target matching. DEM allows achieving excellent integral and differential linearity, with modest component matching. DEM compensates only component variations and not systematic errors (like opamp finite gain, etc.). This also means that DEM can correct only zero-average error; in alternative Calibration technique has to be considered, but they can hardly achieved DEM perfor-

mance. Another important feature of DEM is that it does not require any knowledge of the actual component matching (this is not valid for calibration). This means that DEM is less sensitive to small matching variation due to age/temperature and then it effectiveness is long-term valid. However DEM allows correcting a certain mismatch error until a certain value (to be evaluated with a Montecarlo analysis): beyond that value the performance deteriorate. Among several DEM algorithms, the proper selection depends on the technological capacitor mismatch to be corrected, the application requirements to be achieved, and the digital machine complexity to be implemented. The most popular DEM algorithms: the randomizer, the Clocked Level Averaging (CLA), the Individual Level Averaging (ILA), the Data Weighted Averaging (DWA), the Mismatch Shaping (MS), the tree structure, and some others. The most popular is the DWA, which achieve excellent performance with smaller digital machine complexity. Its only drawback is a certain tone generation that is reduced/canceled by introducing small modifications to the basic DWA algorithm.

4.2 Switched-Capacitor D/A Interface

The Integrating SC D/A interface is shown in Fig. 12. A capacitor (C_s) is precharged to V_{r+} or V_{r-}, depending on the input data and the stored charge ($=C_s \cdot V_{r+}$) is then injected into the virtual ground. The positive point of this structure is that it is insensitive to the clock jitter. In fact, the charge transfer from C_s to C_f occurs at the beginning of the integration phase ϕ_2 and then any eventual ϕ_2 length change does not affect the amount of injected charge. This mechanism guarantees that switched-capacitor circuits offer good coefficient matching and then stable frequency response that can be eventually compensated by the digital filter frequency response. An example of this solution is in [3], where the following performances are achieved: 92 dB-DR, −89 dB-SNDRpeak, with 17 mW from a ±5 V supply.

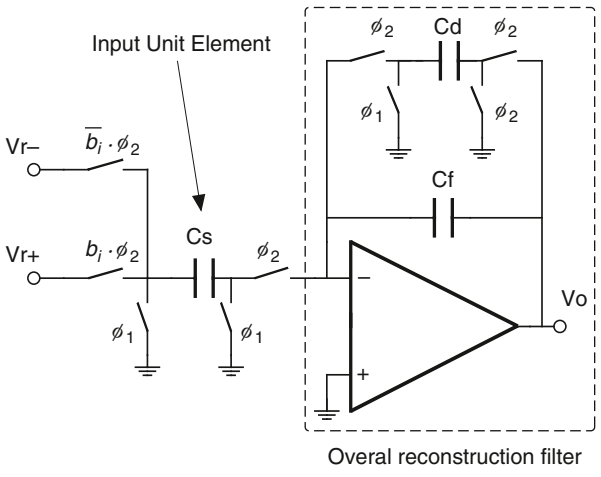


Fig. 12 Integrating SC D/A interface

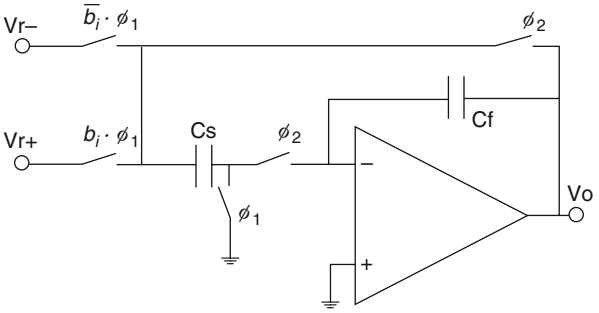


Fig. 13 Deglitcher SC D/A Interface

On the other hand the negative point is that the opamp has to operate with a large sampling frequency (equal to the data-rate) and then with performance and power consumption problems. As already claimed before, analog filter linearity is crucial. Thus in this scheme all the possible solutions for improving linearity have to be adopted. For the switches, 4-phase clock scheme, clock boosting (even if the input signal can assume only two values, with limited linearity problem). The on-resistance is critical for only the switches connected to the output node. Moreover large gain, large slew-rate, large output swing opamp has to be used. Slew-rate is in particular very important. For this reason the deglicher scheme of Fig. 13 can be adopted (called also Direct Charge Transfer—DCT) [4]. This scheme presents some important advantages: no charge transfer is required from the opamp to the integrating capacitor and the noise is not sampled. Both these features allow reducing the opamp power consumption.

In literature several SC implementations demonstrates the validity of this proposal that has been improved by using a multibit structure, an additional input FIR, or introducing a high-order filter.

Figure 14 shows the multibit structure In [5], a 3rd-order ΣΔM produces a 5b data-stream that is processed by a sophisticated 2nd-order DEM (restricted 2nd-order DWA) to address the non-linearity issue. This allows the device to perform −120 dB-SNR, and −100 dB-THD, from a 5 V supply with 200 mW power con-

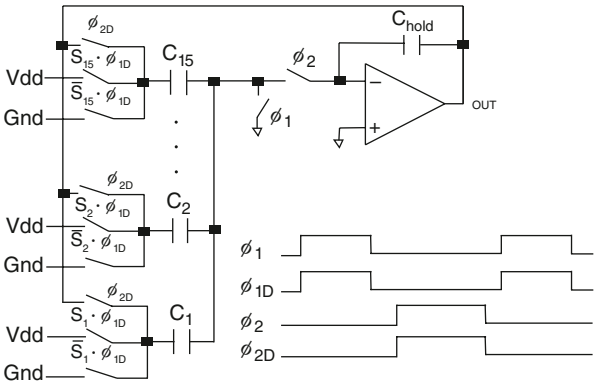


Fig. 14 DCT & multibit structure

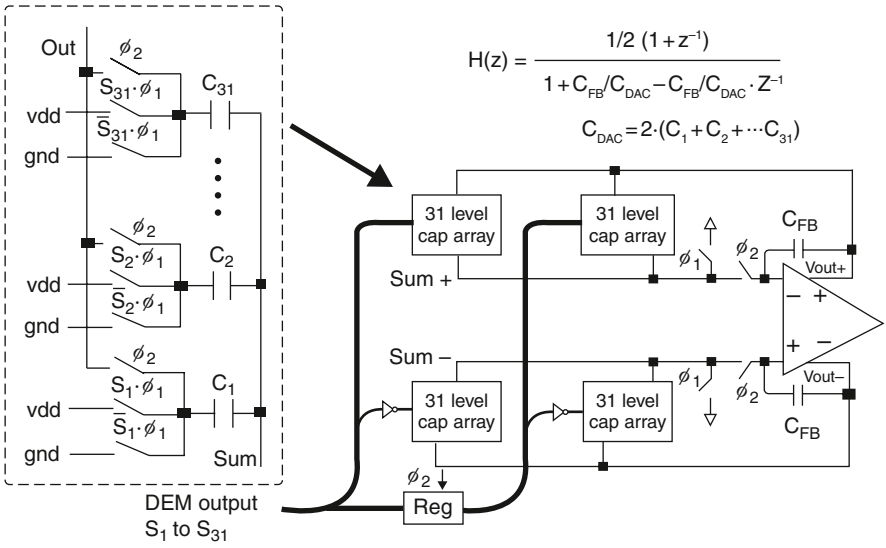


Fig. 15 DCT & FIR structure

sumption (0.35 μm CMOS). In [6] from a reduced 1.5 V supply, the achieved performances are 90 dB-DR, 81 dB-$SNDR_{peak}$ with 4.1 mW.

The DCT coupled with a FIR input structure is shown in Fig. 15 [7]. The achieved performances are 120 dB-DR, 102 dB-SNDRpeak with 310 mW from a 5 V supply.

In Fig. 16 the DCT scheme is coupled with a high-order filter structure [8]. A 3rd-order filter with a pseudo-DCT structure allows a significant reduction of the out-of-band quantization noise. The structure performs also low noise since only the 1st opamp noise reaches the output with unitary gain. This solution performs 98 dB-DR, 86 dB-SNDRpeak with 28 mW from a 3.3 V supply.

A very-compact solution is given in [9]. The digital nature of the DAC input signal is exploited in the D/A interface (Fig. 17) to realized a single-ended structure that avoids the use of fully-differential structures and fully-differential-to-single-ended transformers. Thus this device features small die size and low power consumption. The device achieves a 97 dB-DR and a 39 dB SNR_{out} with a 0.22 mm^2 area and 7.25 mW from a 3.3 V supply (0.13 μm CMOS).

4.3 Switched-Current

The alternative of the SC D/A interface is the switched-current D/A scheme. A passive output network could be propose, with poor driving capability [10]. On the other hand Fig. 18 shows two possible solutions with an opamp output node: one refers to switched-current and the other to switched-resistor. In these schemes in each clock cycle a charge amount given by $Q_{inj} = I_R \cdot T_s$ is injected into the virtual ground

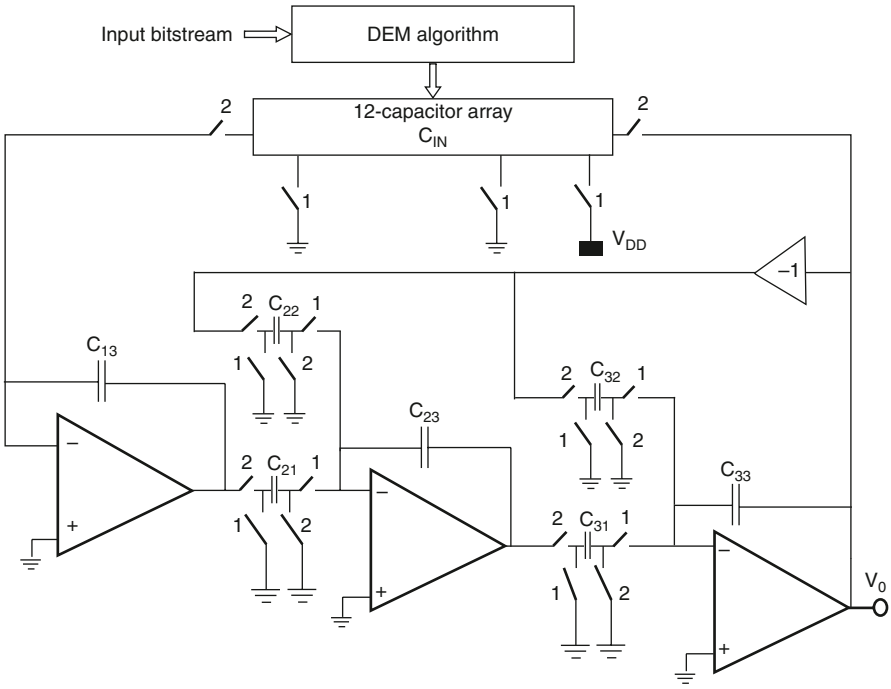


Fig. 16 Pseudo DCT & multibit & high-order filter structure

and integrated into the feedback impedance. The analog filter uses an active-RC structure (2nd-order structure can be implemented), where the specification for the opamp seems to be relaxed with respect to the SC scheme. However the injected charge as shown in Fig. 19 is affected by several circuit implementation imperfections both at the beginning and at the end of the pulse, as described in the following. At the beginning of the pulse, since the input current is a step, the opamp finite speed of response results in a movement of the virtual ground, which modifies the total current injected. For this reason, the opamp bandwidth cannot be so relaxed. In alternative, the current should not depend on the virtual ground voltage, i.e. exhibiting a very large output impedance. Cascode current source are typically used. At the end of the pulse, the negative edge could change due to clock jitter (τ_{jit}) that results in an error in the charge injected into the virtual ground ($Q_{inj}=I_R\,(T_s+\tau_{jit})$).

The error introduced by jitter is very serious and it is strongly limiting current-mode DAC. To overcome this limitation, several solutions have been proposed. A first solution uses properly generated and controlled current injection, adopting, in particular SC techniques [11, 12]. Another solution is using a multibit quantizer: in Fig. 20 the grey area is the error due to jitter that is much lower in a multibit than for a single-bit quantizer. For this reason almost all the current-mode DACs use multibit quantizer. A second solution refers to the consideration that jitter error is introduced in the data transition. Then, limiting data transitions would reduce jit-

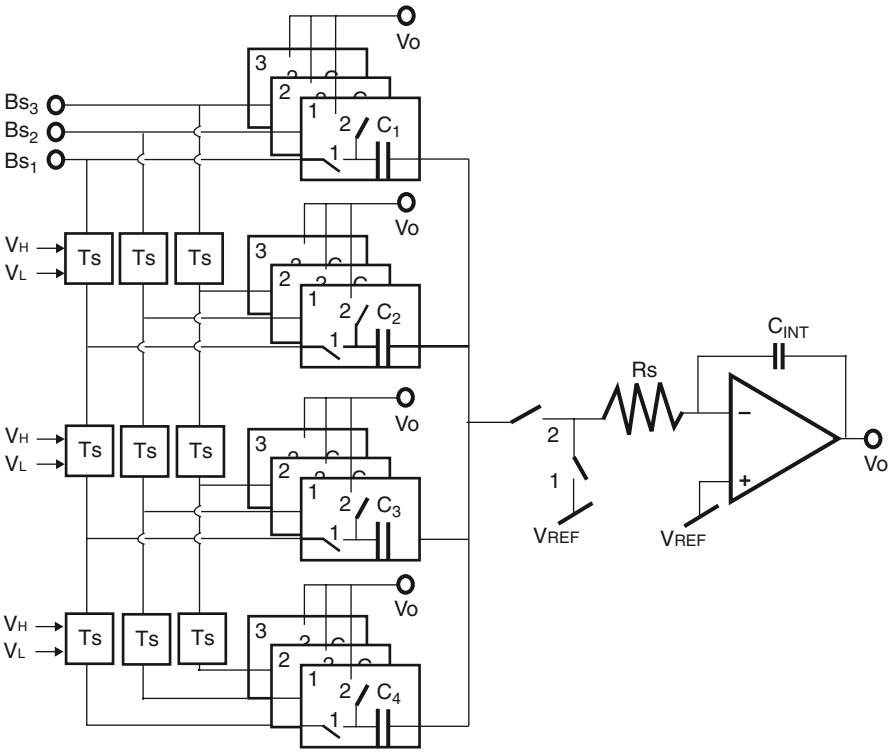


Fig. 17 Single-ended audio-DAC

ter error. This means that two consecutive ones have to be implemented without turning on-and-off the input current, as shown in the Non-Return-to-Zero scheme (NRZ) of Fig. 21.

From another point of view, it can be observed that due to Inter-Symbol Interference (ISI) two consecutive ones gives different total injected charge than two separate ones. Since it is requested that each one contributes with the same inject-

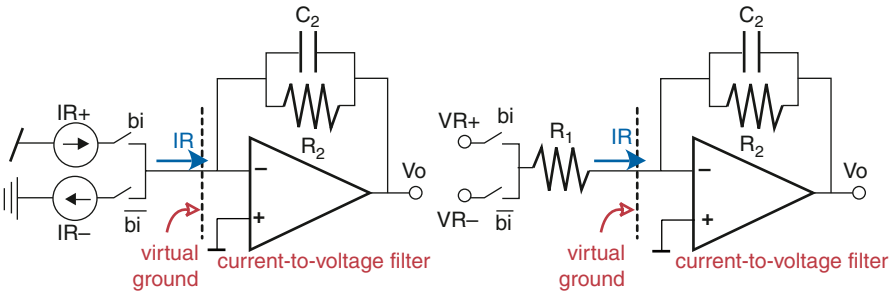


Fig. 18 Switched-current and switched-resistor D/A interface

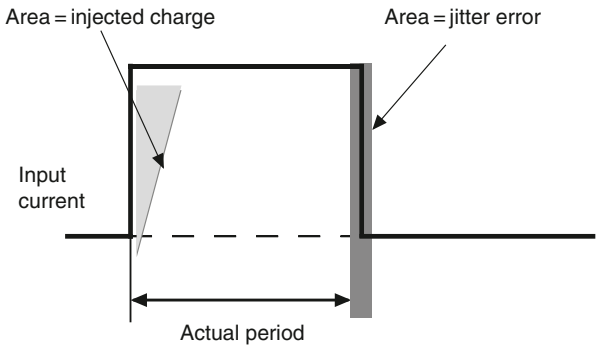


Fig. 19 Jitter sensitivity

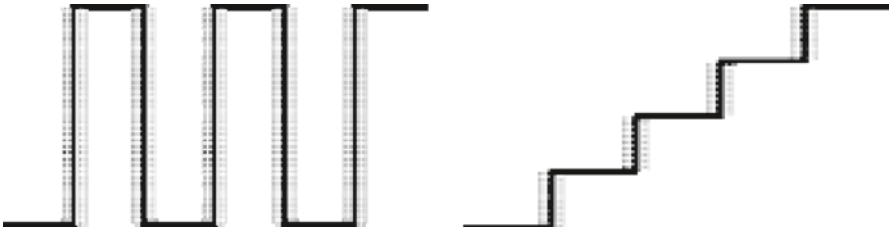


Fig. 20 Single-bit vs. Multibit quantizer for jitter effect limitation

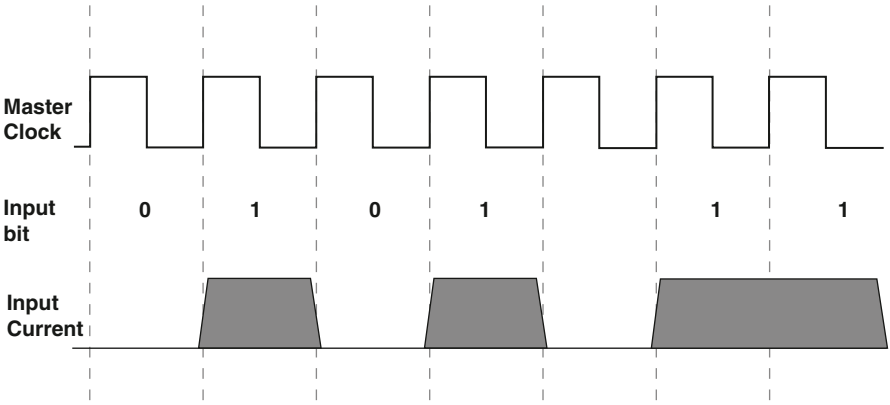


Fig. 21 Non-Return-to-Zero scheme (NRZ)

ed charge, all the consecutive ones have to be separated, as in the Return-to-Zero scheme of Fig. 22. In this case, for each one there is a positive and negative edge and no ISI error occurs. A more sophisticated scheme uses dual-RTZ scheme [2] to achieve (with a 6 bit-ΣΔM) 113 dB-DR, −100 dB-SNDRpeak, with 250 mW.

Notice that there is trade-off between the NRTZ scheme (useful for jitter effect reduction) and the RTZ scheme (useful for ISI effects reduction).

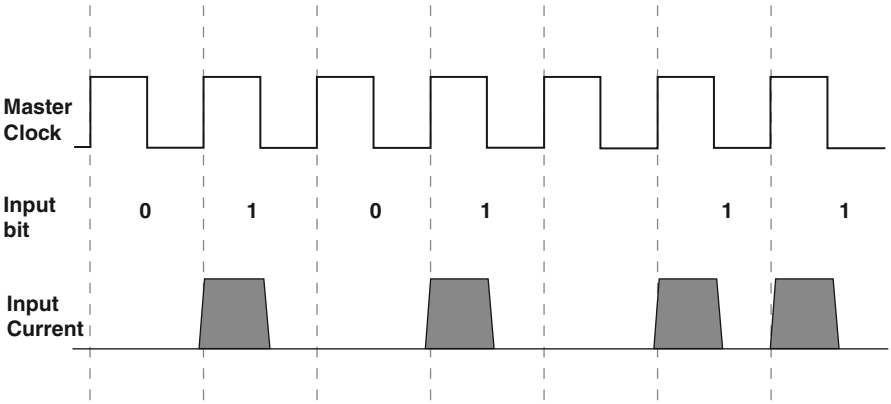


Fig. 22 Return-to-Zero scheme (RTZ)

5 Advanced Oversampled DACs

5.1 Increasing the Quantizer Output Number of Bit

In [13], a 2nd-order 8b-ΣΔM (with OSR = 64) achieves the low out-of-band noise requirement without the need of post-analog filter. This finer modulator resolution also reduces the clock-jitter sensitivity of the output stage and relaxes the application clock requirement. However, the direct use of any DEM technique on the 8b ΣΔM-output would result in an impractical design. For this reason, a noise-shaped segmentation technique [2, 13] is exploited to split the 8b word into 3 segmented outputs, a 4b word and a pair of 3b words with the corresponding weights. A auxiliary 1st-order modulators embedded in the DEM algorithm shapes the DAC-to-DAC gain error into high-passed noise while maintaining the 8b resolution of the main modulator. The 8bit output word allows using a current-steering D/A interface with negligible clock jitter sensitivity. The D/A output current is injected into an Active-RC, which produces the analog output signal. In a 0.18 μm CMOS technology the performance are a 108 dB-DR, and a −97 dB-$SNDR_{peak}$ with 1.1 mW from 1.8 V supply.

5.2 Embedding Power Stage in the Analog Filter

In literature there are several examples in which the power stage is not an external part of the audio channel (as shown in Fig. 23) but is embedded in the DAC reconstruction filter. This solution presents the advantage of compactness, power efficiency, and reduced path for external noise to be coupled in the audio signal path. However it has to face the necessity of operating with a continuous-time out-

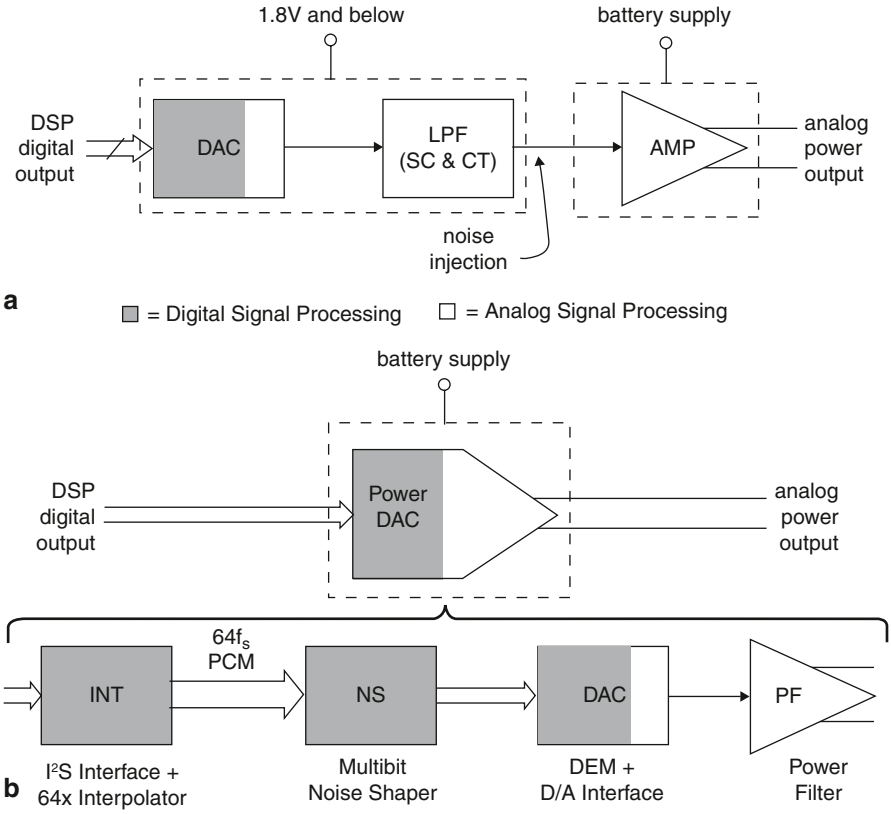
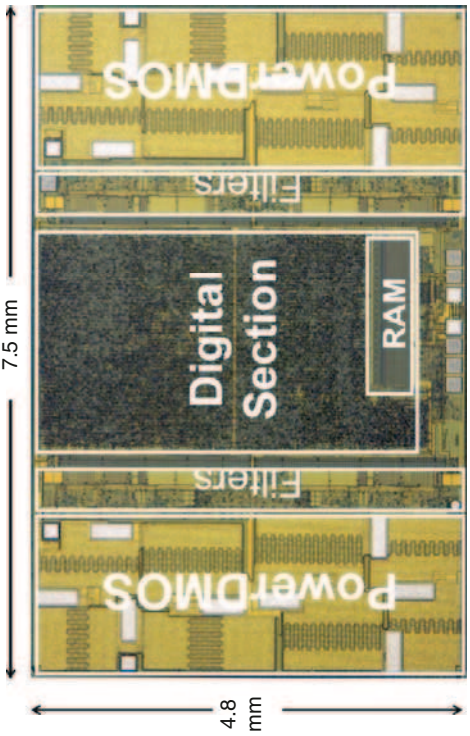


Fig. 23 Output power stage external (**a**) or internal (**b**) in the audio DAC

put stage, eventually operating at higher supply. In addition, sometime the optimum technology for the signal path (short channel device for analog and signal circuits) does not match the optimum technology for the power path (where DMOS are typically preferred). For the output stage in literature several implementations are given. Class-AB could be preferred where no noise has to be generated, disturbing the DAC operation. This however results in reduced power efficiency with respect to other implementation using more sophisticated output stage with higher power efficiency but with poorer signal purity.

One of the first examples of this solutions [14] uses 1b-ΣΔM coupled with a class-D output stage to drive a 8 Ω output impedance. The device delivers 2×1 Wrms with 70% efficiency. The performance includes a 83 dB-DR with a −59 dB-THD for 1 W.

An alternative solution [15] is embedded in a car-radio FM-receiver and uses a class-AB output stage to reduce interference with the RF front-end. It is designed to drive high-power (40 W) and then large output stage is adopted. A 3rd-order, 7b ΣΔM (OSR = 64) feeds a current-mode D/A interface. Active-RC reconstruction filter is used, as shown in Fig. 24. In this scheme, the first opamp operates at low



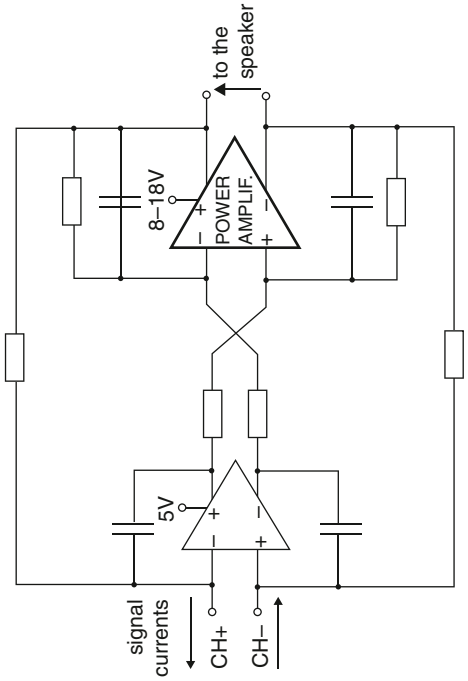


Fig. 24 Power DAC reconstruction filter and layout

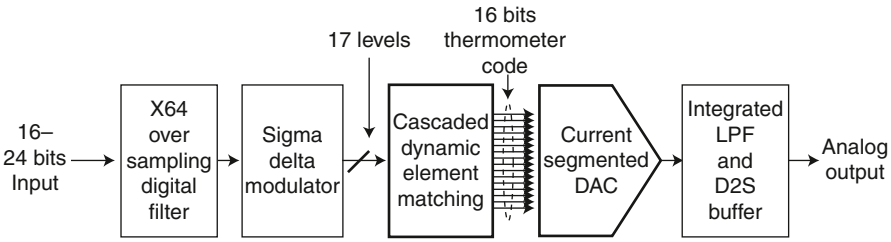


Fig. 25 DAC chain [16]

supply (5 V) and uses "signal" devices to increase bandwidth and, then, reduces virtual ground movement effects. On the other hand the 2nd-opamp operates at high supply (18 V) and uses high voltage devices to properly drive the low-output nodes. The device can deliver up to 40 W and it performs 100 dB-DR.

5.3 Improving PSRR

In [16] the PSRR is optimized by using cascode sources as unit elements of the D/A interfaces. Typically current-sources exhibits lower matching than resistors or capacitors and then DEM has been employed to correct unitary current mismatch (Fig. 25). An improved segment flipping technology with a cascade DEM is introduced to improve matching and jitter tolerance. The device uses a ΣΔM with OSR = 64 and with 17 output levels (the D/A interface is composed by 16 unit elements). The cascaded DEM utilizes 1 master DEM block and several slave DEM blocks. Each DEM block can be simple DWA, high-order mismatch shaping or any of existing DEM. The proposed DAC utilizes 4 slave blocks, thus each slave block takes care of 4 segments (Fig. 26).

Current segment has high tolerance against power supply, however it suffers from two types of mismatch. The one is the mismatch between segments, and the other is the mismatch of differential characteristic. The former mismatch is handled by the cascaded DEM. The segment flipping technology eliminates the latter mismatch. The device performs 106.1 dB-SNR and 0.0018%-THD at full-scale input within 3.39 mm^2 die area, operating from a 5 V supply.

5.4 Cascaded ΣΔM

In [17], a cascade-modulator architecture with parallel DACs that can be weighted arbitrarily to attain high effective resolution levels with greater flexibility is presented in Fig. 27. The error signal across the primary modulator is scaled up by a factor of "K" and passed to the secondary modulator. The scaling down happens on the analog side by sizing secondary elements as "1/Kth" of the primary ones.

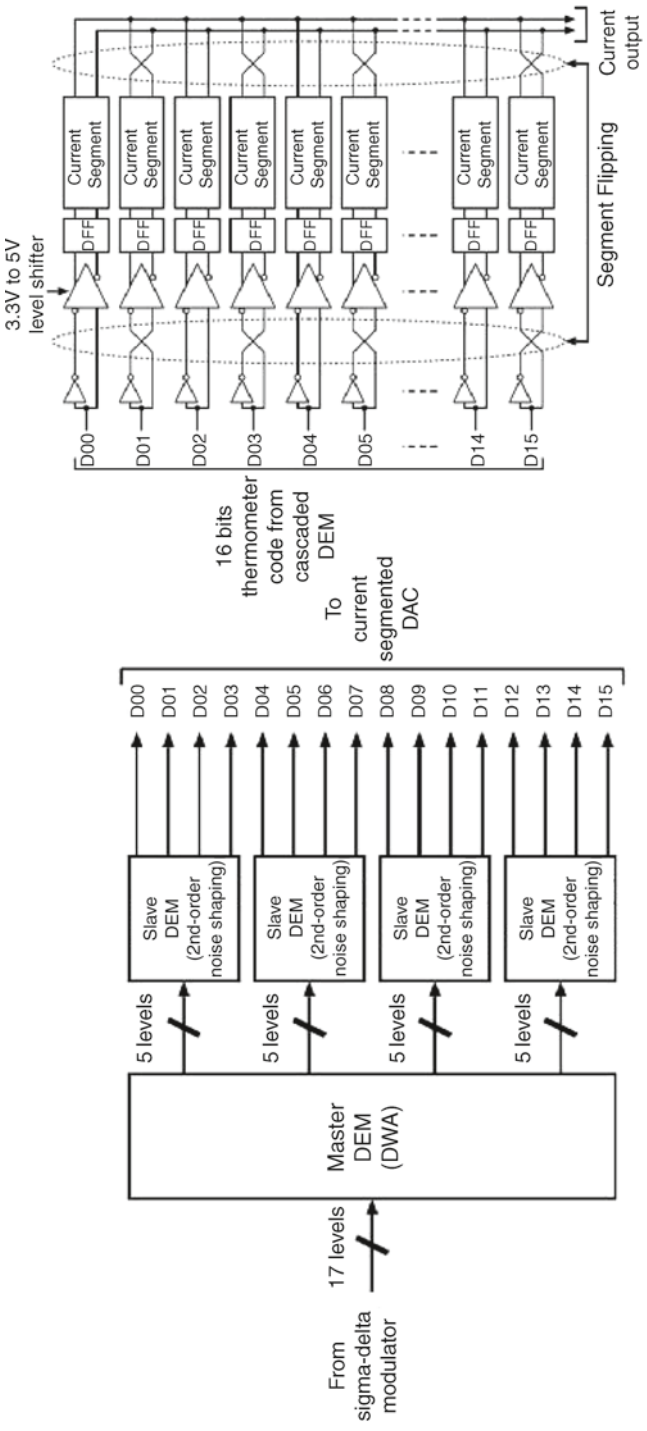

From sigma-delta modulator
17 levels
Master DEM (DWA)
5 levels
Slave DEM (2nd-order noise shaping)
D00 D01 D02 D03 D04 D05 D06 D07 D08 D09 D10 D11 D12 D13 D14 D15
To current segmented DAC
16 bits thermometer code from cascaded DEM
3.3V to 5V level shifter
DFF
Current Segment
Segment Flipping
Current output

Fig. 26 Segmented and flipping DEM implementation [16]

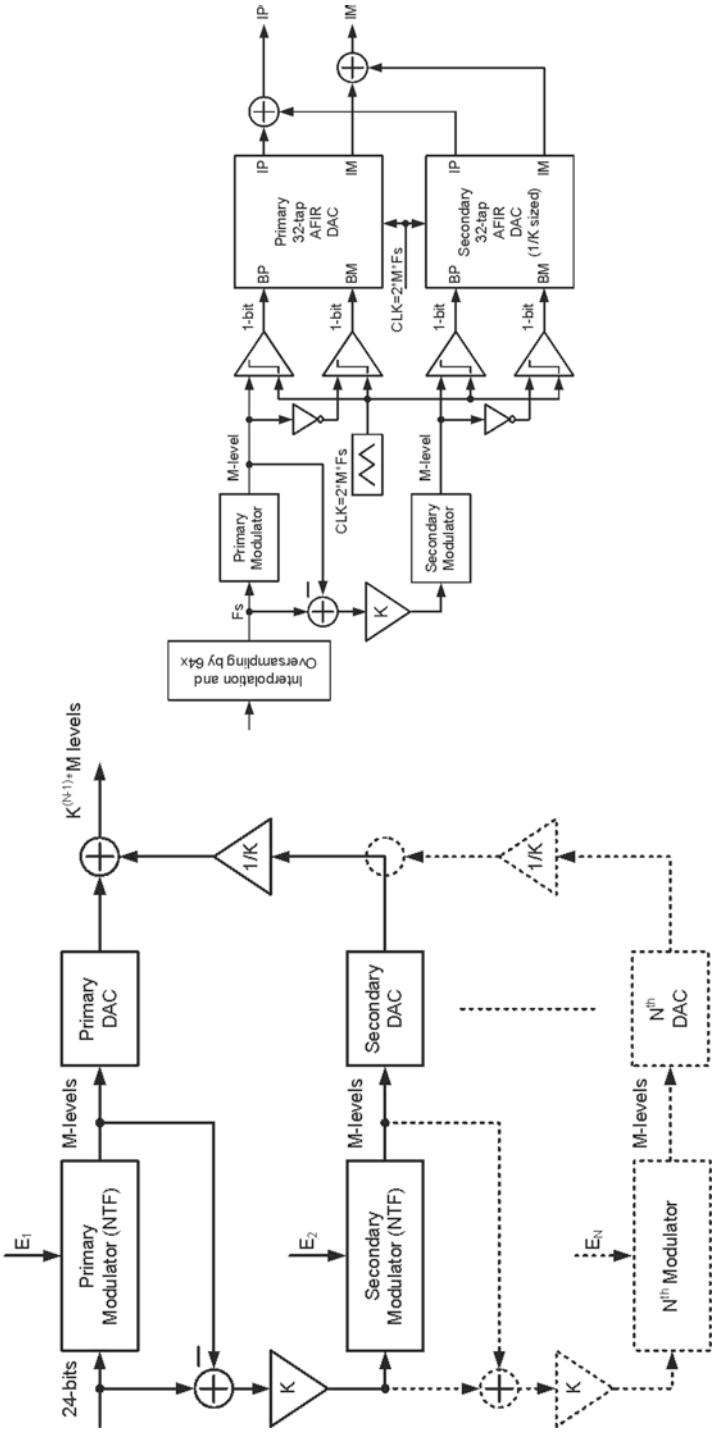


Fig. 27 Cascade ΣΔM-DAC structure and complete DAC architecture with Class-D

Therefore, cost of the secondary path is significantly lower for large values of K. For a simple 2-level cascaded system, the combined output transfer function can be written as

$$Y(z) = X(z) + \frac{1}{K} \cdot NTF(z) \cdot E_2(z)$$

If the secondary modulator has M-levels, then the final output will effectively have K·M level resolution at combined output. Mismatch changes the transfer function as

$$Y(z) = X(z) + \left(1 - \frac{K_d}{K_a}\right) \cdot NTF(z) \cdot E_1(z) + \frac{1}{K_a} \cdot NTF(z) \cdot E_2(z)$$

where K_d, K_a are the digital and analog coefficients, respectively. Any mismatch between the primary and secondary DACs is a deviation in K_a. This mismatch is shaped by the secondary noise transfer function (NTF) and has almost no impact on the in-band signal when compared to the in-band quantization noise contributed by the secondary modulator.

In theory, a 1% mismatch allows to set K as high as 100 before starting to get limited by mismatch. In this architecture, there is no reason that primary and secondary modulators must be identical. Each path can be optimized individually to maximize the overall resolution. The quantization noise of the primary modulator loads the secondary one which is providing finer resolution steps. Therefore, in the overall optimization scheme, one can put the emphasis on reducing the quantization noise when picking the primary modulator, and the stability when picking the secondary modulator. This method also maximizes K-factor, which, in turn, increases the effective resolution. Cascading operation can be further repeated and OBN can be completely buried under the thermal noise floor. This operation is bounded with minimum device size and sensitivity to mismatch within one modulator path. An N-deep cascade with K-scaling at each level will result in $K_{N-1} \cdot M$ levels at the output. Previous noise-shaped segmentation method [1, 2] can still be applied on each modulator path depending on the quantizer levels chosen per path and mismatch shaping operation. The two-path output is coupled with two class-D output stage to improve power efficiency and to reduced ISI-induced noise and distortion in a continuous-time current-mode DAC. The device performs 110 dB-DR, −100 dB-SNDRpeak, with 0.5 mW in 45 nm technology.

5.5 Improved ΣΔM for Idle-Tone Reduction

In [18] the ΣΔM is modified in order to reduce the idle tone energy that is typically generated when low-amplitude signal is applied to the ΣΔM. The scheme if the modified SDM is shown in Fig. 28, where the reference levels of the digital multibit ΣΔM (OSR = 128) are modulated by a shaped version of the quantization noise. In

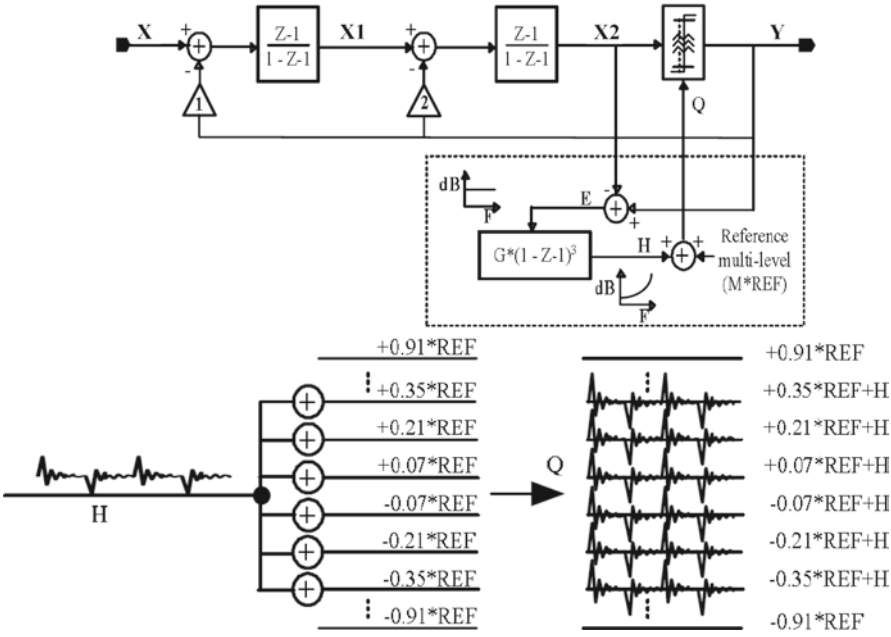


Fig. 28 Modified ΣΔM for idle-tone reduction

addition, an improved version of the DWA is applied to the 16-level SC D/A interface. The reconstruction filter is the cascade of a DCT SC stage and of a two-pole active-RC filter.

5.6 Hybrid DAC

In literature there are also some hybrid DACs that use both SC technique (mainly for the D/A interface) and continuous-time active-RC (for the filtering stage) [19].

A detailed example of this kind of hybrid oversampled DAC with embedded a power stage to drive an external load (a loudspeaker) is here presented. Embedding the power amplifier allows to save one opamp and so area and power consumption. Moreover the power amplifier is embedded in the filter loop and then its requirements in terms of distortion and noise are relaxed. On the other hand, the hybrid structure exploits the switched capacitor technique for the input stage and the feedback DAC in order to gain un-sensitivity versus clock jitter (thus increasing performance in terms of THD and noise) and to have more flexibility in the design of input/output common modes. As a last observation, the case of a load dc-connected with the output stage is considered. This solution avoids decoupling capacitor but suffers from an eventual offset that would means power continuously delivered to

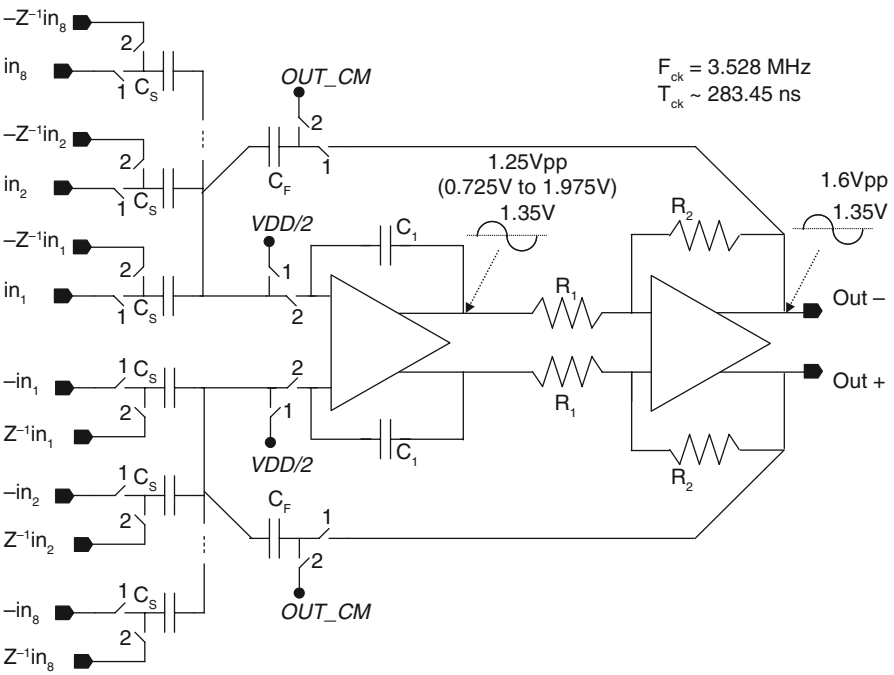


Fig. 29 Hybrid SC-continuous time reconstruction filter structure with power amplifier

the load, with a consequent important limitation of the battery duration. For this reason a new technique to reduce the offset contribution realizing a feedback at the input of the digital modulator is here proposed.

Figure 29 shows the reconstruction filter architecture: a 2nd-order fully differential hybrid SC/continuous-time structure, in which the Power Amplifier (PA) acts as a second stage inside the loop. Input digital signal is injected inside the first stage virtual ground through 8 SC branches. To obtain a good compromise between high in-band SQNR (>100 dB) and low out-of-band noise SQNR (>50 dB), a low-order ΣΔM multi-bit structure and a DEM is used in the multi-bit D/A SC interface. The adopted SC D/A interface allows the possibility to easy realize a $(1+z^{-1})$ filter that gives two benefits: it reduces the out-of-band quantization noise and it realizes a 2× gain in the audio band allowing a reduction in the capacitor sizing. In the biquad cell scheme of Fig. 29, the 2nd integrator exploits the PA pole instead of using a capacitor in parallel to R_2. This results in an additional power (the PA bandwidth can be smaller) and area (the cap area is avoided) saving. The most important design parameters are collected in the following table.

The PA bandwidth should be low enough in order not to become un-stable when the inductive load is connected: speakerphones, earphones, headsets available on the market have an inductive component in the range of tenth of micro-Henry. To

Technique	Hybrid, Switches Capacitor/Continuous Time
Clock frequency (Fs)	3.528 MHz
Bandwidth of interest	20 Hz–20 kHz
OSR	88.2
Input branches	28
Output voltage Diff.	3.2 Vpp,diff
Rec. Filter Architecture	Chebychev, 2nd order
Open loop poles	0 (Integrator) 80 kHz (PA with its resistive feedback)
Closed loop poles	~70 kHz
Static current consumption	<600 μA

be compliant with a generic actuator, PA bandwidth should be lower than 100 kHz. In this condition, it is not possible to realize the second filter pole (80 kHz) with a capacitor in parallel to R_2 (a 3rd pole introduced by the PA should be too close and it would drive to un-stability). Thus, the PA pole has been exploited to realize the second reconstruction filter pole. This is the most important draw-back for this kind of architecture, because the settling of the switch in feedback is limited by the load if the output impedance of the PA is not low enough at clock frequency. This draw-back has to be carefully evaluated and simulated in order not to introduce distortion. Figure 30 shows the Reconstruction Filter signal transfer function: the slope variation at 2 MHz is due to the SC nature of the first integrator. This integrator has 20 dB/dec only for frequencies far from Fs/2 (3.528/2 MHz). This effect marginally impacts only on the out-of-band noise rejection.

The adopted structure has a THD strongly dependent on the input frequency. Considering that the SC 1st-Integrator can guarantee a THD up to −100 dB the PA results to be the main THD contributor. The THD dependency with the input frequency is 40 dB/dec. First 20 dB is due to the Power Amplifier Gain: it decreases by 20 dB/dec in the band of interest, according to Fig. 31. A second 20 dB/dec is due to the SC 1st-Integrator attenuation that decreases by 20 dB as the frequency decreases.

Figure 32 collects simulation results on THD at various input gains and signal frequencies, showing the 40 dB/dec. gain as the input frequency decreases.

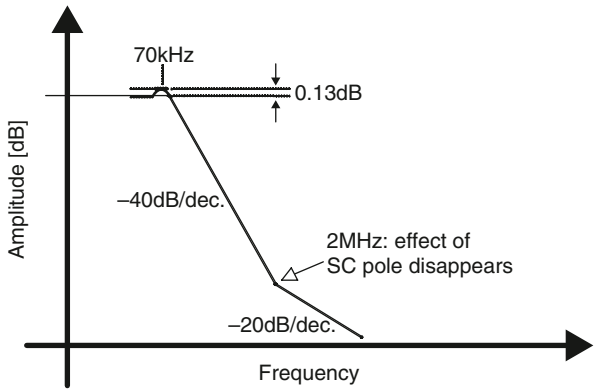


Fig. 30 DAC filter frequency response

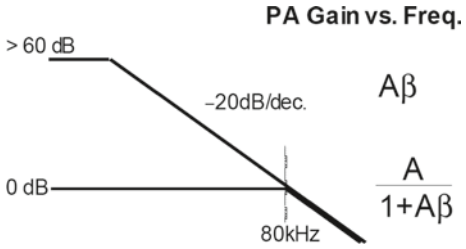


Fig. 31 THD vs. frequency effects

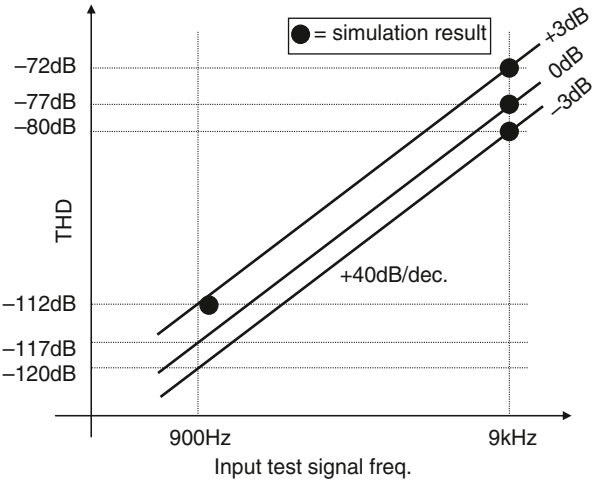


Fig. 32 THD vs. input frequency (extrapolated values, starting from simulation results)

The noise analysis of the devices is carried out evaluating each noise contribution with separate simulations as given in the following table.

Two key problems reduce the performance of this device. In a DC-coupled differential output stage, the presence of an offset at the PA output requires larger

	Simulated/calculated value [V^2]	Transfer function	Over-sampling effect	Power @ output [V^2]
Signal Power				$3.2^2/2=5.12$
Power @ load				1.28 W (4 Ω)
Quantization noise	55.6 p	–	–	80 p
Opamp Flicker noise	49 p	$(1+\Sigma C_S/C_F)^2$	1	237 p
Opamp Thermal noise	9 n	$(1+\Sigma C_S/C_F)^2$	1/OSR	496 p
kT/C sampling caps	$2\cdot2\cdot14.4$ p	$(\Sigma C_S/C_F)^2$	1/OSR	536 p
kT/C feedback caps	$2\cdot2\cdot8.6$ p	1	1/OSR	447 p
Not filtered PA Flicker noise	–	–	–	67 p
Not filtered PA Thermal noise	–	–	–	46 p
R1/R2 noise	–	–	–	9.5 p
Total				1843 p
DR				94.3 dB

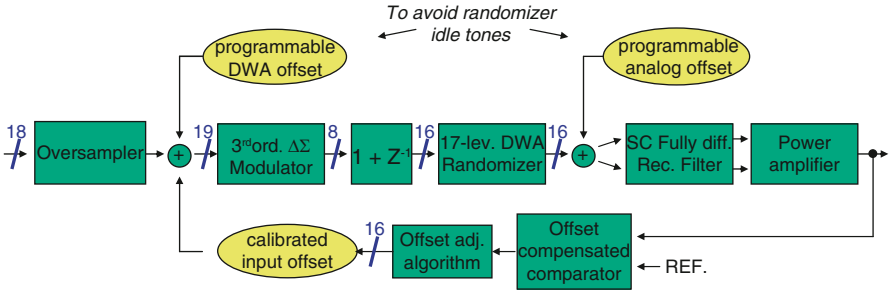


Fig. 33 DAC block diagram with offset compensation circuit

power consumption. The DEM DWA algorithm [3] for an odd number of levels results in the same sequence for low input signal amplitude. This produces some idle tones and, then, reduces the DEM algorithm effectiveness.

Figure 33 describes a closed-loop control system acting to compensate the offset due to analog blocks. An eventual offset at the PA output is detected by a reading circuit, which is realized by means of a narrow-band offset-compensated comparator. Its output is fed to the following offset adjustment algorithm (that can be simply realized as an up & down counter, i.e. a digital low-pass filter) that tends to reduce it adding a "calibrated input offset" at the Sigma-Delta-Modulator input.

In addition to the word for the offset compensation, another DC-contribution (of variable amplitude) is added. This additional contribution disables the DWA algorithm to enter in the limit cycle, which increases harmonics generation. This additional offset contribution has to be present only at the DWA level, while it would be dangerous in the analogue blocks since it could reduce the available swing, introduce distortion, and reduce the final dynamic-range.

For this reason, this offset amount is subtracted at the multi-bit digital-to-analogue interface by means of a dedicated unitary element added to the standard unitary elements array. In this way this additional offset does not enter into the analogue part.

This solution can be further optimized, canceling the additional DWA offset at the SDM input. In this case, the loop controls the cancellation of the offset amount at the PA output. If some offset is added at the D/A interface, this offset is compensated by the calibrated input offset of the servo loop. In this way, the servo loop is producing the additional offset previously specifically introduced as "programmable DWA offset". The amplitude of this offset is now controlled by the "programmable analogue offset" amplitude.

This second approach is simpler and, so, more efficient. Actuating the offset control in front of the SDM modulator is due to the fact that this is the point with the maximum available accuracy (18 bit). This resolution corresponds to the accuracy of the offset cancellation at the PA output. This accuracy is also affected by the "Offset Adjustment Algorithm": this is a successive approximation algorithm requiring an iteration number equal to the resolution bit number. Assuming a counter (low-pass filter) the length of the counter window affects its accuracy and then there

will be a trade-off between a short algorithm convergence time and an accurate offset cancellation (achieved for a long algorithm convergence time).

The proposed device in a 0.15 μm CMOS technology features a 94 dB-DR (with an 8 Ω load) with a 76 dB-SNRpeak. The static power consumption is 600 μA.

6 High-Speed Oversampled DAC

The concept described in the previous chapters have been extended to higher signal bandwidth applications as soon as the scaled technologies give the possibility of using correspondingly higher sampling frequencies for the oversampled ΣΔM. In these cases the key choices are forced by the limited oversampling frequency that can be used in the digital part, in order not to exceed the power consumption. Thus a multibit D/A interface is used, based on switched-current with Active-RC filter (which requires a lower UGB opamp). The DEM algorithm is typically a DWA, which performs sufficient mismatch correction at a very small cost in term of area and power consumption.

In [20], a 1.1 MHz signal bandwidth (for ADSL Central Office Modems) is processed with on OSR = 100, i.e. with a sampling frequency of 200 MHz. The device achieves 86 dB-DR and 71 dB-$SNDR_{peak}$, with 82 mW from a 1.8 V supply (0.18 μmCMOS technology). For the same application in [21], using a lower OSR in a 0.18 μm technology, with a 35.3 MHz clock frequency, the performances are 73.9 dB-$SNDR_{peak}$, 79.8 dB-SFDR with 55 mW from 1.8 V supply. Other examples of these extended applications are given in [22–24], where using a switched-current DAC interface a 29.16 MHz signal bandwidth is processed with $F_s = 350$ MHz (i.e. with OSR = 6). The structure presents also an active transimpedance output stage to provide output voltage. The device performs 73.4 dB-DR, 76 dB-THD, and consumes 45 mW fro single 1.5 V supply in 0.13 μmCMOS.

References

1. H.J. Schouwenaars, D.W.J. Groeneveld, C.A.A. Bastiaansen, H.A.H. Termeer, An oversampled multibit CMOS D/A converter for digital dynamic audio with 115-dB range. IEEE J. Solid-State Circuits. **26**(12), 1775–1780 (1991)
2. R. Adams, K. Nguyen, K. Sweetland, A 113-dB SNR oversampling DAC with segmented noise-shaped scrambling. IEEE J. Solid-State Circuits. **33**(12), 1871–1878 (1998)
3. J. Peicheng, K. Suyama, P.F. Ferguson Jr., L. Wai, A 22-kHz multibit switched-capacitor sigma-delta D/A converter with 92 dB dynamic range. IEEE J. Solid-State Circuits. **30**(12), 1316–1325 (1995)
4. N. Sooch, High order switched-capacity filter with DAC input. US-Patent #5,245,344
5. X.-M. Gong, E. Gaalaas, M. Alexander, D. Hester, E. Walburger, J. Bian, A 120 dB multi-bit SC audio DAC with second-order noise shaping, in *ISSCC*, 2000, pp. 344, 345, 469
6. I. Fujimori, T. Sugimoto, A 1.5-V 4.1-mW dual-channel audio delta–sigma D/A converter. IEEE J. Solid-State Circuits. **33**, 1879–1886 (1998)

7. I. Fujimori, A. Nogi, T. Sugimoto, A multibit delta-sigma audio DAC with 120 dB dynamic range. IEEE J. Solid-State Circuits. **35**(8), 1066–1073 (2000)
8. M. Annovazzi, V. Colonna, G. Gandolfi, F. Stefani, A. Baschirotto, A low-power 98 dB multibit audio DAC in a standard 3.3 V 0.35 µm CMOS technology. IEEE J. Solid-State Circuits. **37**(7), 825–834 (2002)
9. V. Colonna, M. Annovazzi, G. Boarin, G. Gandolfi, F. Stefani, A. Baschirotto, A 0.22 mm^2 7.25 mW per-channel audio stereo-DAC with 97 dB-DR and 39 dB-SNR_{out}. IEEE J. Solid-State Circuits. **40**(7), 1491–1498 (2005)
10. T. Hamasaki, Y. Shinohara, H. Terasawa, K.-I. Ochiai, M. Hiraoka, H. Kanayama, A 3-V, 22-mW multibit current-mode DAC with 100 dB dynamic range. IEEE J. Solid-State Circuits. **31**(12), 1888–1894 (1996)
11. M.-Y. Choi, S.-N. Lee, S.-B. You, W.-S. Yeum, H.-J. Park, J.-W. Kim, H.-S. Lee, A 101-dB SNR hybrid delta-sigma audio ADC using post integration time control, in *CICC*, 2008, pp. 89–92
12. M. Ortmanns, Y. Manoli, F. Gerfers, A continuous-time sigma-delta modulator with reduced jitter sensitivity, in *ESSCIRC*, 2002, pp. 287–290
13. K. Nguyen, A. Bandyopadhyay, B. Adams, K. Sweetland, P. Baginski, A 108 dB SNR 1.1 mW oversampling audio DAC with a three-level DEM technique, in *ISSCC*, 2008, & IEEE J. Solid-State Circuits. **43**(12), 2592–2600 (2008)
14. K. Philips, J. van den Homberg, C. Dijkmans, Power DAC: a single-chip audio DAC with a 70%-efficient power stage in 0.5 µm CMOS, in *ISSCC*, 1999, pp. 154, 155
15. C. Meroni et al., A 100 dB 4 Ω 40 W digital input class-AB power DAC for audio applications, in *IEEE Int. Solid-State Circuits Conf. (ISSCC) Dig. Tech. Papers*, Feb. 2003, p. 130
16. T. Ido, S. Ishizuka, T. Hamasaki, A 106 dB audio digital-to-analog converter employing segment flipping technology combined with cascaded dynamic element matching, in *VLSI*, 2005, pp. 174, 175
17. R. Hezar, L. Risbo, H. Kiper, M. Fares, B. Haroun, G. Burra, G. Gomez, A 110 dB SNR and 0.5 mW current-steering audio DAC implemented in 45 nm CMOS, in *ISSCC*, 2010, pp. 304, 305
18. Y.-H. Lee, M.-Y. Choi, S.-B. You, W.-S. Yeum, H.-J. Park, J.-W. Kim, A 4 mW per-channel 101 dB-DR stereo audio DAC with transformed quantization structure, in *CICC*, 2006, pp. 145–148
19. B.M.J. Kup, E.C. Dijkmans, P.J.A. Naus, J. Sneep, A bit-stream digital-to-analog converter with 18-b resolution. IEEE J. Solid-State Circuits. **26**(12), 1757–1763 (1991)
20. A.C.Y. Lin, D.K. Su, R.K. Hester, B.A. Wooley, A CMOS oversampled DAC with multi-bit semi-digital filtering and boosted subcarrier SNR for ADSL central office modems. IEEE J. Solid-State Circuits. **41**(4), 868–875 (2006)
21. P.A. Francese, P. Ferrat, Q. Huang, A 13b 1.1-MHz oversampled DAC with semidigital reconstruction filtering. IEEE J. Solid-State Circuits. **39**(12), 2098–2106 (2004)
22. M. Clara, W. Klatzer, A. Wiesbauer, D. Straeussnigg, A 350 MHz low-OSR ΔΣ current-steering DAC with active termination in 0.13 µm CMOS, in *ISSCC*, 2005, pp. 118, 119
23. D. Giotta, P. Pessl, M. Clara, W. Klatzer, R. Gaggl, Low-power 14-bit current steering DAC, for ADSL2+/CO applications in 0.13 µm CMOS, in *ESSCIRC*, 2004, pp. 163–166
24. P. Seddighrad, A. Ravi, M. Sajadieh, H. Lakdawala, K. Soumyanath, A 3.6 GHz, 16 mW ΣΔ DAC for a 802.11n/802.16e transmitter with 30 dB digital power control in 90 nm CMOS, in *ESSCIRC*, 2008, pp. 202–205

Part III
RFID

Although RFID is surely not new, it is attracting more attention in recent years. One reason for that is the take off in the market. RFIDs are really becoming more and more popular, which of course is due to the fact that tag technology has advanced now so far, that price, size and performance have reached levels that made market penetration possible. But, technology for making the tags itself, is not all; companies want solutions integrated in their processes, and that too takes time and requires new technologies, but then in the sense of assembling and manufacturing technologies. And, obviously, also standards were (and will be) very crucial for mass adoption.

Aside these evolutionary trends, there is another reason for the increased interest: new technologies, that can open completely new markets and applications. Most importantly, organic electronics, that provides an alternative to the silicon-based technologies, and printing technologies, that (often, but not necessarily, in combination with organic electronics) promise very cheap mass production for flexible products.

This section on RFID starts with a contribution of Henri Barthel, from the global office of the standardization institute GS1 in Brussels. It puts the RFID tags into a context, seen from different angles: historically, markets, applications, standardization, privacy issues, and regulations.

The next three papers address tags made in silicon IC technology. The first one, of Mitsuo Usami from Hitachi, describes the design of the world's smallest, and thus also very low cost, silicon-based tag IC. Aside the IC-design aspects it treats novel fabrication and assembling technologies that enable low-cost fabrication. In a next chapter, Raymond Barnett of Texas Instruments addresses analysis and design aspects for the various blocks of complex tags that require full functionality. The third chapter, of Albert Missoni from Graz University and Infineon, describes the design of a front end for a multi-mode tag that can operate at both the UHF and HF RFID frequency band.

Finally we end up with two papers describing first transponder blocks made in emerging new technologies. The first one, from Jürgen Krumm, PolyIC, focuses

on printed electronics. The second one, from Kris Myny of IMEC/Holst Centre, focuses on organic electronics. Both describe the state of the art in these fields and the expectations for the future.

Arthur H. M. van Roermund

RFID, a Technology Ready for Industry Deployment

Henri Barthel

1 RFID History

Radio frequency identification has been around for decades. It's generally said that the roots of radio frequency identification technology can be traced back to World War II. The Germans, Japanese, Americans and British were all using radar to warn of approaching planes while they were still miles away. The problem was there was no way to identify which planes belonged to the enemy and which were a country's own pilots returning from a mission. The Germans discovered that if pilots rolled their planes as they returned to base, it would change the radio signal reflected back. This crude method alerted the radar crew on the ground that these were German planes and not Allied aircraft (this is, essentially, the first passive RFID system) [1]. The British developed the first active identify friend or foe (IFF) system. They put a transmitter on each British plane. When it received signals from radar stations on the ground, it began broadcasting a signal back that identified the aircraft as friendly. RFID works on this same basic concept. A signal is sent to a transponder, which wakes up and either reflects back a signal (passive system) or broadcasts a signal (active system).

In the 1970s, Los Alamos National Laboratory in the U.S. was asked by the Energy Department to develop a system for tracking nuclear materials. A group of scientists came up with the concept of putting a transponder in a truck and readers at the gates of secure facilities. The gate antenna would wake up the transponder in the truck, which would respond with an ID and potentially other data, such as the driver's ID. This system was commercialized in the mid-1980s when the Los Alamos scientists who worked on the project left to form a company to develop automated toll payment systems. These systems have become widely used on roads, bridges and tunnels around the world.

At the request of the Agricultural Department, Los Alamos also developed a passive RFID tag to track cows. The problem was that cows were being given hormones and medicines when they were ill. But it was hard to make sure each cow got the right dosage and wasn't given two doses accidentally. Los Alamos came up with a passive

H. Barthel (✉)
GS1 Global Office, Brussels, Belgium

H. Casier et al. (eds.), *Analog Circuit Design,*
DOI 10.1007/978-94-007-0391-9_13, © Springer Science+Business Media B.V. 2011

RFID system that used UHF radio waves. The device drew energy from the reader and simply reflected back a modulated signal to the reader using a technique known as backscatter. Later, companies developed a low-frequency (125 kHz) system, featuring smaller transponders. A transponder encapsulated in glass could be injected under the cow's skin. This system is still used in cows around the world today. Low-frequency transponders were also put in cards and used to control the access to buildings.

Over time, companies commercialized 125 kHz systems and then moved up the radio spectrum to high frequency (13.56 MHz), which was unregulated and unused in most parts of the world. High frequency offered greater range and faster data transfer rates. Companies, particularly those in Europe, began using it to track reusable containers and other assets. Today, 13.56 MHz RFID systems are used for access control, payment systems and contactless smart cards. They're also used as an anti-theft device in cars. A reader in the steering column reads the passive RFID tag in the plastic housing around the key. If it doesn't get the ID number it is programmed to look for, the car won't start.

In the early 1990s, IBM engineers developed and patented an ultra-high frequency (UHF) RFID system. UHF offered longer read range and faster data transfer. IBM did some early pilots with Wal-Mart, but never commercialized this technology. When it ran into financial trouble in the mid-1990s, IBM sold its patents to Intermec, a bar code systems provider. Intermec RFID systems have been installed in numerous different applications, from warehouse tracking to farming. But the technology was expensive at the time due to the low volume of sales and the lack of open, international standards.

UHF RFID got a boost in 1999, when the Uniform Code Council, EAN International, Procter & Gamble and Gillette put up funding to establish the Auto-ID Center at the Massachusetts Institute of Technology. Two professors there, David Brock and Sanjay Sarma, had been doing some research into the possibility of putting low-cost RFID tags on all products made to track them through the supply chain. Their idea was to put only a serial number on the tag to keep the price down (a simple microchip that stored very little information would be less expensive to produce than a more complex chip with more memory). Data associated with the serial number on the tag would be stored in a database that would be accessible over the Internet.

Sarma and Brock essentially changed the way people thought about RFID in the supply chain. Previously, tags were a mobile database that carried information about the product or container they were on with them as they travelled. Sarma and Brock turned RFID into a networking technology by linking objects to the Internet through the tag. For businesses, this was an important change, because now a manufacturer could automatically let a business partner know when a shipment was leaving the dock at a manufacturing facility or warehouse, and a retailer could automatically let the manufacturer know when the goods arrived.

Between 1999 and 2003, the Auto-ID Center gained the support of more than 100 large end-user companies, plus the U.S. Department of Defense and many key RFID vendors. It opened research labs in Australia, the United Kingdom, Switzerland, Japan and China. It developed two air interface protocols (Class 1 and Class 0), the Electronic Product Code (EPC) numbering scheme, and a network architec-

ture for looking up data associated on an RFID tag on the Internet. The technology was licensed to the Uniform Code Council in 2003, and the Uniform Code Council created EPCglobal, as a joint venture with EAN International, to commercialize EPC technology. The Auto-ID Center closed its doors in October 2003, and its research responsibilities were passed on to Auto-ID Labs. In 2005, the Uniform Code Council became member of EAN International and the new organisation was named GS1. The 108 national EAN organisations were renamed GS1 followed by the name of the country, e.g. GS1 US, GS1 France.

2 RFID Market Landscape

The BRIDGE project, discussed in detail in Sect. 13.4, commissioned a detailed market analysis on passive RFID, which was published early 2007. Based on this analysis we predicted that by 2012 more than 170,000 passive RFID readers will be deployed in Europe at 30,000 locations. These readers will process a total of 3 billion tags. These numbers will grow significantly until 2022, when we expect more than 6 million readers to be operating at 450,000 locations, with 86 billion tags purchased annually.

We believe these numbers to be conservative, as they only represent a small percentage of the total potential number of objects that can be tagged. For example, our forecast is based on the estimation that in 2012 2% of all items in retail will be tagged. In 2022 we forecast that roughly 25% of all non-food items and 5% of all food items in retail will be tagged. If we experience a technology breakthrough in the next fifteen years that would reduce the cost of an RFID tag to less than one cent, these number could increase dramatically. In particular the number of tags on food items could grow to hundreds of Billions.

Other key findings of this research:

- High-value item-tagging will remain the largest opportunity for RFID tag and reader volumes. In the short term we expect to see significant growth in fashion and apparel, cultural goods (DVD's, books etc) and consumer electronics. In these categories RFID is helping to improve inventory management in the store, which may result in less stock outs.
- For many RFID applications it will probably take another two-three years before the market will really take off. This is due to technical issues, price levels that are prohibitive for the business case, or discussions on the distribution of costs and benefits in open supply chains.
- In the long term we expect hardware costs to come down dramatically. This is due to a combination of technical innovations and economies of scale. A passive RFID reader may cost 200 €, and tag prices may come down to a few Euro cents. A potential breakthrough in chipless technologies may even result in tag prices of less than 1 € cent (although our forecasts are more conservative and not based on this assumption).

	2007	2012	2017	2022
Total number of tags purchased annually (in Millions)	144	3.220	22.400	86.700
Total number of locations with RFID readers	2.750	30.710	144.000	453.000
Total number of RFID readers deployed	7.630	176.280	1.161.800	6.268.500

Fig. 1 RFID market sizing 2007–2022 [2]

- Retail & consumer goods will remain the largest market in terms of volumes for RFID tags and readers, accounting for approximately two-thirds of the total market volume, both short term and long term.
- After retail, the postal & express market provides the most potential. In the short term this market will focus on areas such as returnable transport items, but longer term it is likely to adopt RFID in its core processes on parcels and mail.
- In aviation RFID will be used for a wide range of applications. In 2012 we expect the first large-scale implementations in Europe in baggage tracking, and this will continue to grow in the future.
- In the next five years we do not expect RFID to be widely used in Europe against counterfeiting of drugs. Instead, we expect the pharmaceutical industry to focus on 2D barcode implementations in the near future. More longer term, the industry may change to RFID.
- UHF will be the dominant frequency for the tracking of physical objects, with HF being used in a number of niche markets such as library books.

In conclusion, RFID is developing very differently from the way most people envisioned a few years ago. And in five years from now we may again discover that RFID adoption has not developed as we predicted in this study. However, this uncertainty only relates to the speed of adoption across the various RFID applications.

There is no doubt that performance will continue to improve and prices will continue to come down. It seems certain that in the future passive RFID will become an integral part of doing business in Europe, resulting in Millions of readers that read Billions of tags each year (Fig. 1) [2].

3 RFID Standards

3.1 The Core Architecture

RFID standards are already published or being developed to cover aspects of an RFID system from the tag through to data exchange with business partners. Figure 2 shows a comprehensive RFID system architecture [3].

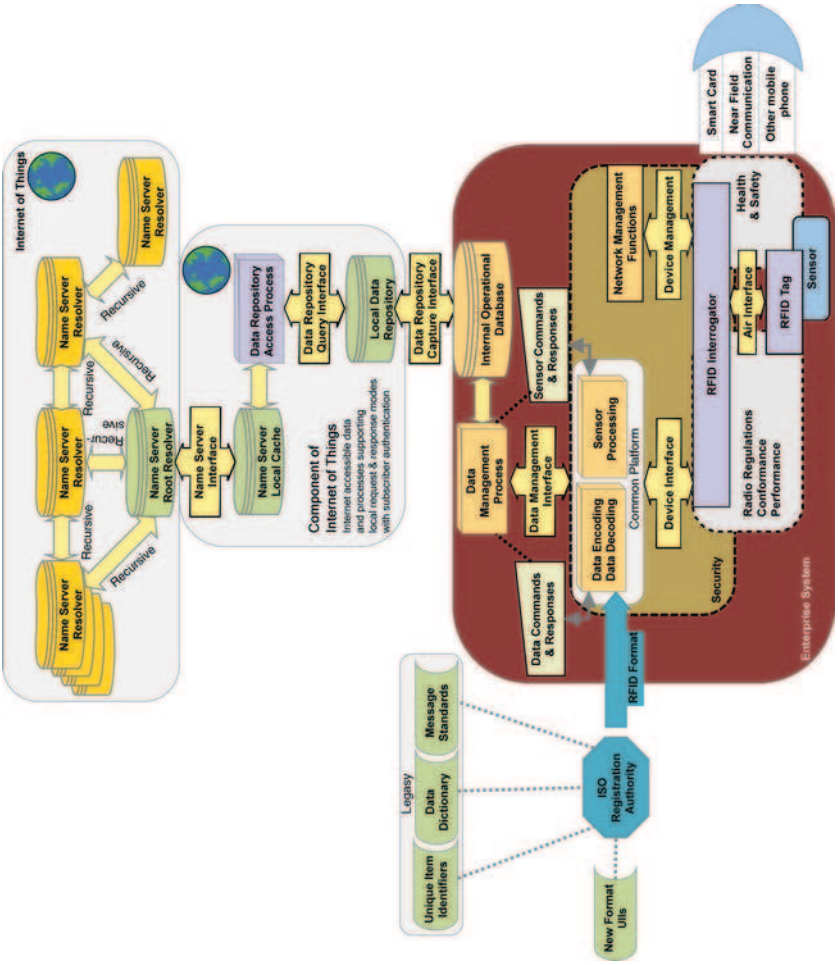


Fig. 2 RFID system architecture [3]

Some of the standards in the architecture are still in a development phase. No application currently uses all the components, even of standards that have been published.

We have sub-divided the architecture into four component parts:

- The enterprise system, dealing with all aspects where RFID as a data carrier is used to assist with some functional aspects of business or commercial operations.
- Internet-based data exchange components that are internal to the enterprise.
- Internet-based data exchange that is external with partners and other stakeholders.
- The ISO Registration Authority for data format that provides support for converting legacy data and for new forms of unique item identifiers.

3.2 The Formation of ISO/IEC JTC1 SC31 WG4

The first proposal for the development of a standard in the field of RFID in ISO/IEC JTC 1/SC 31 Automatic Identification and Data Capture Techniques was presented on 13 March 1997. The scope of the new work item proposal was 'To define a communication protocol for interoperability of wireless, non-contact omnidirectional radio frequency identification devices capable of receiving, storing, and transmitting data while operating at power levels that are in freely available international frequency bands in the area of item level identification and management across the supply chain such as finished good asset management, raw material asset management, material traceability, inventory control, electronic article surveillance, warranty data, production control/robotics, and facilities management.'

Following the acceptance of this new work item proposal SC31 Working Group 4 (WG4) was created. SC31 WG4 had its first meeting on 26–28 August 1998 in Tokyo, Japan.

3.3 RFID Standards and Applications Before the Epoch

Before the formation of SC31 WG4, a number of initiatives to standardise RFID and related technology had taken place. Within ISO itself, the work of JTC1 SC17 on smart card technology was well established and this represented an eventual overlap at 13.56 MHz with the developments in JTC1 SC31 WG4 RFID for Item Management. The United States of America had a number of standards in place under the ANSI/INCITS committee work that were potential candidates for ISO standardisation.

ISO, and other organisations, had undertaken major work developing RFID technology associated with specific applications. These included:

- RFID technology and application standards for animal ID.
- European and ISO standards using RFID for road traffic telematics.
- A set of UPU work-in-progress standards specifying RFID technology at each of the main frequencies.

Apart from the US ANSI standards, there was little previous work on developing standards that only identified the technology and covered a range of RFID frequencies. Most of the candidate technologies were initially considered by WG4 as proprietary. Because the technology was proprietary, most of the applications—including some that would eventually migrate to open systems—were for closed systems often limited in scale. These covered a great variety of niche applications—some could be considered to be forerunners of RFID in the supply chain. There were applications for all the main RFID frequencies, with the exception that UHF technology was not used for RFID in many parts of the world because of the potential overlap of that frequency being used for mobile telephones.

3.4 Bar Code Applications and Application Standards Before the Epoch

Bar code and RFID had similar development paths, and were first invented at a similar time in the decade to 1950. The development of bar code, like RFID, was based on proprietary technology, and the companies developing the technology provided solutions for closed system applications.

The major transition for bar code came in the early to mid 1970s. A keynote development was the formation of the Uniform Code Council and then, within a short period of time, EAN International (renamed GS1 in 2005). Other significant organisational developments were taking place in parallel. The manufacturers of bar code technology created their trade association, Automatic Identification Manufacturers (AIM) originally as a product section of the Materials Handling Institute in 1972. The objective was to develop technical solutions, and promote the technology. A significant number of technology exchanges took place, which resulted in previously proprietary bar code symbologies becoming de facto standards. These were finally formalised in 1982 by the publication of a set of symbology specifications. By 1983, AIM had become an independent trade association in the United States, and shortly afterwards progressed internationally with the formation of AIM UK and AIM Europe.

Developments continued, and members of AIM worked with many different sectors concerned with the development of application standards. Another milestone was the creation of the data dictionary for primary industry originally known as FACT Data Identifiers and later adopted and standardised by ANSI.

By the time of what we have called the "epoch", the formation of SC31 WG4 RFID for Item Management, SC31 WG1, WG2 and WG3 had been in place dealing predominantly with bar code. CEN TC225 had been in existence for some time before this.

The net effect was that by the epoch the bar code vendor community had a deep understanding of application requirements, and the user organisations responsible for developing applications either had significant technical understanding of bar code technology, or could tap into industry resources for this.

Depending on one's perspective and analysis, the progress of RFID was between 12 and 25 years behind what had been achieved with bar code technology by the epoch.

3.5 Early RFID Standardisation Activities

As previously highlighted, there were some fundamental differences in SC31 with the process to standardise bar code and to standardise RFID. By the time SC31 was created, a number of bar code symbologies (data carriers) had already been published as standards by AIM (which is a formal standards committee of ANSI), by CEN for European standards, and other National Bodies. Therefore, the initial task for bar code was to address the quality of the standard not the technology. Over time, more bar code symbologies were submitted to SC31, but often a significant amount of prior work had been done elsewhere.

In contrast, RFID activities began with almost a blank agenda, although by the time the two RFID Ad Hoc meetings had taken place, there were some clear directions. For a long time, the work of the air interface protocol and the application interface protocol were seen as distinctly separate entities, although there were experts with crossover experience. The brief for the application interface activity was to develop encoding rules that were to be largely independent of air interface issues.

With the benefit of retrospective analysis, the air interface standardisation activity seems to have progressed through three phases. The first phase was concerned with structuring rules between and within the standards. Whereas each bar code symbology is a distinctly separate data carrier technology standard, an early decision made for RFID was that each part of ISO/IEC 18000 would address a particular frequency. The ideal expressed at the time was to have one air interface protocol per frequency, which was based on a model borrowed from the Universal Postal Union's work on RFID. However, there were many candidate technologies, most of which were effectively proprietary technologies.

The second phase was concerned with the process of reducing nearly 30 candidates into the limited number of standards. For quite some time, there was no consensus on which technology was best in class. So, every time technologies were being evaluated and compared, there were vested interests among other experts, which usually resulted in more informal votes against any given technology than in favour. A key meeting took place in Marseille, France where all the technology sponsors were offered a straightforward ultimatum. Either face perpetual ballots with the end result of no RFID standards, or work formally or informally together

to reduce the variety of candidates. This did lead to a significant amount of collaboration outside the scope of ISO, and enabled the ISO/IEC 18000 standards to move forward and meet the new, more realistic, criteria of up to two air interface protocols per frequency.

At this time, there was still a lot of enthusiasm based on market expectations, and many of the meetings had between 40 and 60 experts participating. It was inevitable that new technology solutions would emerge. Objective criteria that had been applied to bar code standardisation since the formation of SC31 were adapted for RFID. Basically, any new air interface protocol had to be significantly different from those already standardised or carry with it the support of major applications. On this basis, the first serious attempts were made to achieve global acceptance of the UHF frequency with the result of ISO/IEC 18000-6 being developed and published with Type A and B protocols.

The next most significant development, which involved crossover activities between ISO experts, the Auto-ID Labs and eventually EPCglobal, was the process that resulted in the EPCglobal Class 1 Gen 2 (ISO/IEC 18000-6 Type C) air interface. The current position, when the revisions are published, is that:

- four air interface protocols will be included in ISO/IEC 18000-6 (UHF technology): Type A, Type B, Type C, and Type D (previously known as TOTAL).
- three in ISO/IEC 18000-3 (high frequency): Mode 1, Mode 2, and Mode 3

3.6 ISO/IEC JTC1 SC31 WG4

The title of ISO/IEC JTC1 SC31 WG4 is ‘RFID for Item Management’. Since its formation, this working group has published 22 standards and technical reports. At this moment 9 of the WG4 documents are being revised and 13 new documents are being developed.

SC31/WG4 now has the following structure:

- Subgroup 1—Application interface protocols
- Subgroup 3—Air Interface
- Subgroup 5—Implementation Guidelines
- Subgroup 6—RFID Performance and Conformance test methods

4 The BRIDGE Project

4.1 Project Overview

BRIDGE (Building Radio Frequency IDentification for the Global Environment) was developed in response to the call for projects released in May 2005 by the

European Union Information Society Technologies (IST) Directorate under the specific programme “Integrating and strengthening the European research area” in the Community sixth framework programme (FP6).

BRIDGE was an integrated project that started in July 2006 and ended in August 2009. Its consortium was composed of 31 partners, including 7 GS1 organisations, 5 universities, 11 solution providers and 8 user companies. GS1 Global Office coordinated the project.

The objective of the BRIDGE project was to research, develop and implement tools to enable the deployment of RFID and EPCglobal Network applications. The project has developed easy-to-use technological solutions for the European business community, including Small & Medium sized Enterprises (SMEs), ensuring a basis for collaborative EPCglobal systems for efficient, effective and secure supply chains.

The project consisted of a series of business, technical development and horizontal activities. Seven Business Work Packages (WPs) were set up to identify the opportunities, establish the business cases and perform trials and implementations in various sectors including anti-counterfeiting, pharmaceuticals, textile, manufacturing, re-usable assets, products in service and retail non-food items. The project included an important research and development program in various aspects of RFID hardware, software, network and security. A series of horizontal activities provided training and dissemination services, enabling the adoption of the technology on a large scale in Europe for the sectors addressed by BRIDGE and beyond.

The chart below shows the three main work areas, by Work Package group, in the BRIDGE project (Fig. 3).

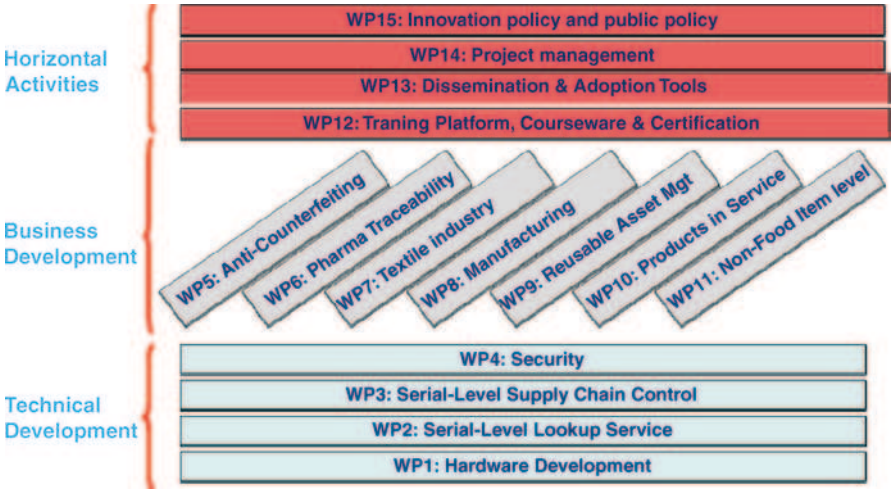


Fig. 3 The three main work areas, by Work Package group, in the BRIDGE project

Technical Development Four working groups concentrated on RFID Hardware, serial lookup services, serial supply chain control and Security. Main achievements included:

- The **Hardware** team developed new RFID tags (more versatile, equipped with sensors, smaller and cheaper, more adaptable for metal and dielectric objects), new RFID readers and reader antennas (less expensive, more performing), new RFID systems prototyping some smart object environments.
- **Serial level Lookup Service** developed the requirement analysis and technical design documents for a Discovery Service, which has greatly contributed to the standards development on Discovery Services. This group also developed a software prototype implementing the original Discovery Service concept.
- **Serial level Supply chain control** developed a tracing model based on track and trace probabilistic algorithms.
- **Security** issued a comprehensive security analysis documenting the requirements for enabling open and collaborative RFID-based business applications. They focused on security and privacy, RFID system security and integrity and network infrastructure security and developed several prototypes.

Business Development The seven working groups looked at business developments in the use of RFID in a variety of sectors. Their work was organised following a common pattern (Fig. 4):

Most of the pilots conducted under the different application work packages started during the second year of the project and were finalised in the third and final year. Many have brought impressive results and interesting outcomes in terms of lessons learnt and best practices. These include:

- **In anti-counterfeiting**—development of new services in the EPCglobal network that will reduce the level of piracy of goods, which is a serious problem in Europe and worldwide,
- **In pharmaceuticals**—increasing patient safety by improving traceability, and certifying the traceability of pharmaceutical products as they move from the manufacturer to the hospital/pharmacy,

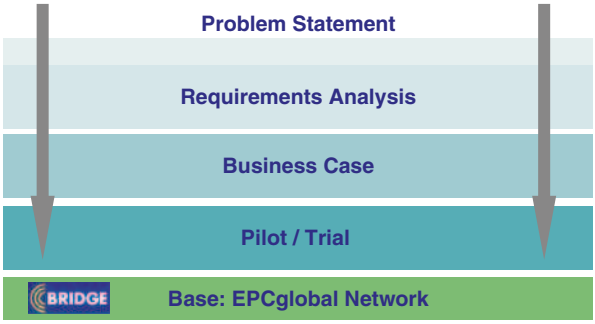


Fig. 4 Business applications development in the BRIDGE project

- **In the textile industry**—better fulfilment of customers needs by increasing the flow and accuracy of information through the supply chain and in the store,
- **In food manufacturing processes**—reducing waste and stock holding and improving visibility and traceability of products and equipment, thereby improving food safety,
- **In re-useable assets**—improving information exchange and asset management between supply chain partners to effect reduction in losses and costs,
- **In products in-service**—developing systems and processes to increase the reliability of the upgrade, repair and replacement processes throughout the life of many products,
- **In the retail environment**—optimising processes in retail stores in order to increase service to the customer by using RFID on consumer sale units.

Horizontal Activities In the first year of the project, the project produced concept animations (multimedia learning objects that illustrate the use of RFID/EPC and their applications) to support training requirements. An RFID market sizing study providing forecasts of the number of readers, tags and locations that would use RFID technology in 5, 10 and 15 years was released early 2007. In the second year of the project, a portable demo that shows how EPC/RFID works in real supply chain conditions was finalised and is now downloadable on the public website http://www.bridge-project.eu.

BRIDGE has also developed five high-level training courses covering basic to expert level, has undertaken the translation in five languages of an awareness raising website on RFID http://discoverrfid.org and undertook vast analysis in regards to data protection issues.

This project has provided a great opportunity for Europe to build on a standardised RFID technology for use in global supply chains. The BRIDGE project has clearly contributed to the development of new solutions for all businesses, both small to large. Improving skills and expertise on RFID technology and network information sharing is leading to enhanced competitiveness of European companies and to benefits to customers and citizens.

4.2 BRIDGE Hardware Developments

BRIDGE worked to advance the state of the art of RFID hardware. The goal was to develop:

- New RFID tags:
  - More versatile, equipped with sensors
  - Smaller and cheaper
  - More adaptable, for metal and dielectric objects
- New RFID readers and reader antennas:
  - Less expensive
  - Better performance
- New RFID systems to prototype some smart object environments

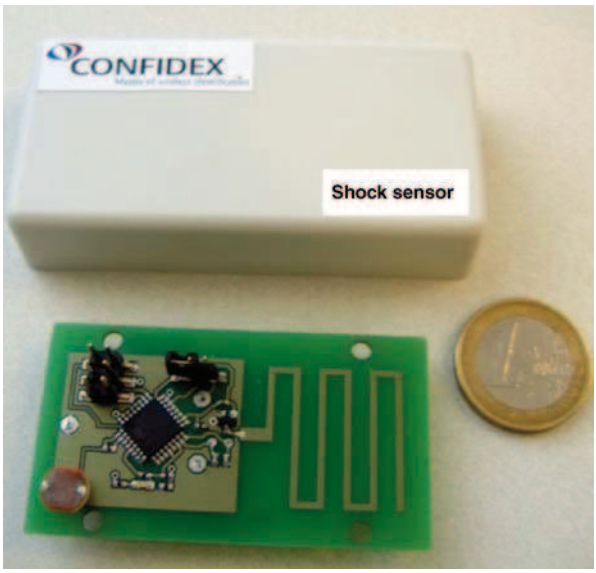


Fig. 5 Prototype of a sensor-enabled tag, compatible with the EPC Gen2 protocol

In the field of *sensor-enabled tags*, first a benchmark study was performed investigating the different technologies, standards, and user requirements. This was compiled into a Sensor-enabled RFID Handbook that reduces the learning curve for any company seeking to develop this kind of tags. Later a design effort was started that resulted in a common platform, proposing operation modes, data management procedures, and protocol extensions to build modular sensor-enabled RFID tags. Finally, some multi-sensor tag prototypes were built based on the proposed common platform (Fig. 5).

In response to a high market demand for *smaller tags* several techniques where investigated for tag miniaturization, from fractal shapes to use of different materials. Finally, a tag with a very high read-range to size was designed, borrowing a concept from meta-materials research: the Split-Ring Resonator (SRR). Tags for *metal and dielectric material* were designed using the same principle: isolation from the material but with a minimum thickness to keep the tags conformal to the shape of the tagged objects. This has been a very competitive field of research, in which industry has come up with high performance designs. The contribution of BRIDGE was a very thin design based on a double bow-tie resonator.

Two different research lines were started to design a *low-cost RFID reader*. On the one hand, the first prototypes in the industry based on specific RFID reader chipsets were designed and prototyped. The forecast was a factor-of-five cost reduction versus current market prices. On the other hand, an effort was started to reduce the price of the RFID chipset itself by designing it using common CMOS processes, rather than using a different process for the RF and the digital parts of the chip. This promises cost reductions of at least a factor-of-ten.

Research was also conducted to improve the *performance of the readers*. For this BRIDGE research concentrated on the reader antennas. First a novel design of

a phased array antenna was tested, improving the read rates of static constellations of tags. It is well known that moving tags are easier to read since multi-path cancellation blind-spots tend to be static. When tags are static, the phased array reader antenna moves the beam slightly, and randomly, moving the multi-path blind spots, and thereby increasing the readability of large constellations of static tags.

Next another difficult RFID problem was tackled: the *metallic shelf*. The smart-shelf concept, in which the shelf is aware of its contents, is an old RFID paradigm that has faced difficulties in practice since most shelves in use today are metallic and no reliable, robust, and economic way had been found to equip such shelves with RFID antennas. To address this, a specific design based on slot antennas was prototyped and successfully tested in a live retail environment, producing 100% read rates over a full month of testing. In this test, multiplexing was used to reduce the total cost of equipping a supermarket with RFID antennas used on metal shelves (Fig. 6).

Finally, a more theoretical research was conducted, with the goal of pushing the limit of how many tags a reader can read in a second. This number is limited by the

Fig. 6 RFID-equipped metal shelf in its test environment

anti-collision protocols. Since all tags communicate with the reader using the same protocol, the reader can only communicate with one tag at a time, and needs to employ a multiple-access strategy, similar to those employed in other one-to-many networks, to resolve collisions. The use of Blind Signal Separation (BSS) algorithms was shown to allow a single reader to communicate simultaneously with up to four tags. For this, readers would have to be equipped with at least four RF front-ends. Each of these receives a different mix of the signals from the four tags. The role of the BSS algorithm is to separate the response from each individual tag. Once this is done, the reader can also communicate back to the four tags simultaneously.

The third area of research demonstrated the use of RFID for *building smart-object systems*. First, the smart-shelf prototype was equipped with algorithms to use the RFID antennas to manage a stock of books in a store, locating them precisely on the shelf, sending out-of-stock alerts, and producing lists of misplaced items. Second, the smart-object paradigm was applied to the remote servicing of heavy equipment. For this a lab prototype of a washing machine was built, which made use of RFID readers and sensors to detect malfunction (over-heating, vibration, water leaks), misuse (wrong washing program for the clothes in the load), use of non-original spare parts, as well as sending warnings when specific parts needed servicing or replacement. A web-based platform was developed to remotely control all of these functions, demonstrating how RFID could dramatically improve the quality and efficiency of the management of a large and geographically disperse fleet of machines (washing machines, vending machines, vehicles, agricultural and mining equipment, etc.).

In summary, BRIDGE research has helped advance the state-of-the art of RFID hardware with a set of prototypes ready for industrialization; promising concepts that require further research in the laboratory; and with some theoretical results. Some of these deliverables will be patented and some will be exploited commercially.

5 Privacy Concerns

In general, RFID is not yet a widely accepted technology, as it raises concerns of privacy and security. A few years ago these led to protests campaigns against early adopters in apparel and retail trade. Although these events now seem to be beyond the shoulders, these concerns were confirmed during the European RFID public consultation (http://www.rfidconsultation.eu) of 2008 [4].

Therefore, privacy and security issues will influence adoption of RFID technology and also imply the need for development of secure RFID, which in turn could make the design, production and deployment of RFID more complex, and more expensive.

Passive RFID tags, like those attached to items, can be almost invisible. It means that people can carry them with object they posses without being aware of this. This implies different kinds of risks. First, RFID tags attached to the object that people bought can be interrogated by someone and reveal what items a person has in a bag (including for example medicines) or what was the price of tagged clothes. More-

over, although the set of things which a person carries is changing, it usually does not change completely. Such a set, called RFID shadow or RFID constellation of a person, if regularly updated, may serve the effective tracking of individuals. This has raised concerns among privacy organizations and individuals.

Some consumers are afraid of function creep, i.e. using a large amount of data legally obtained by an RFID system for different purposes than the original objective of the systems.

The basic security measure against unauthorized reading of RFID tags attached to items is deactivation of the tag at supermarket check-out.

An important economic implication of privacy and security issues is the need to follow technical and legal measures which make RFID (single tags as well as whole systems) more complex, and therefore more expensive. At the level of tags, some measures are already mandatory under EPCglobal standards. However, there is a limit for security of tags: those used for item-level tagging are mostly passive and are not supposed to perform complex functions.

The demand for privacy can be seen as a market opportunity. Apart from the demand for security built into RFID systems, we can foresee demand for personal devices which help the user keep control over tags he owns.

6 Regulations

RFID operates in the electro-magnetic spectrum that is highly regulated all over the world. The most common RFID technologies use Low Frequency (LF, below 135 KHz), High Frequency (HF, 13,56 MHz) and Ultra High Frequency (UHF, 860–960 MHz). The regulations for using RFID at LF and HF are largely available worldwide.

RFID at UHF represents a fast growing market. Huge progress has been made over the last 10 years. In 2010, it is safe to assume that countries representing 96% of global national income have adequate regulations for using RFID at UHF.

In Europe, RFID at UHF operates in the band 865–868 MHz. Four 200 KHz wide channel are used by readers with a power of up to 2 W e.r.p. The other channels are used by tags to respond to the readers (Fig. 7) [5].

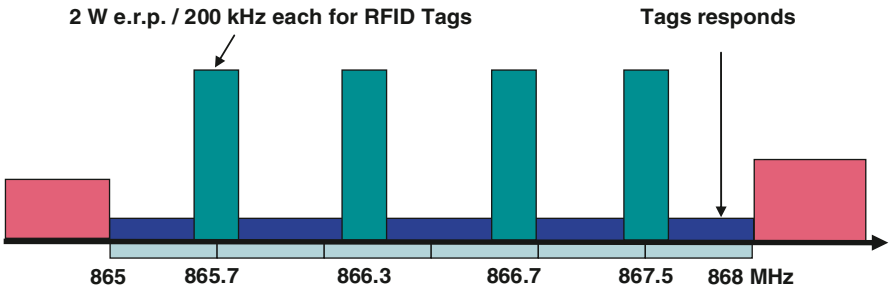


Fig. 7 4 channel plan for RFID at UHF in Europe [5]

The expected need for high usage densities, higher data speeds, greater ranges, faster reading of increased numbers of tags, will require additional spectrum. To satisfy the above requirements it will be necessary to designate more spectrum that will provide high performance reader channels. An additional frequency designation in the range 915–921 MHz is proposed for high performance RFID systems. This will harmonize the operation of RFID in Europe with many of the other major non-European countries that generally operate in the 900–930 MHz range.

According to a recent survey, the logistics' industry and RFID system providers have stated that:

- The density of RFID tags in a typical load or transport unit can be up to 1,000–1,500 tagged objects. The separation between tags typically can be between 2–3 mm to 30 cm.
- According to the application the minimum separation between fixed interrogators as well as hand/mobile readers can be between 0 and 1 m. This means that multiple channels are required in order to avoid tag to tag cross-interference.
- The reading range should be up to 9 m.
- The time available for reading all tags in a bulk/pallet environment varies from 0.5 to 2 s max. Also the data carried by a tag can vary from 96 to 256 bits.
- Acceptable reading performance on a pallet should approach 100%. This requires higher power as well as high data rates for increased read redundancy.
- Additional features (e.g. monitoring of temperature, shock, humidity, pressure) will be used in critical applications to assure the condition of delivered goods. This may require higher power and much increased data capacity.
- The performance of European RFID systems and tags should match global performance. Presently this cannot be guaranteed since European frequencies and power levels do not provide comparable reading performance to e.g. US, SA, Far East, Canada and South American countries.

7 Conclusion

RFID is a fascinating technology. It offers opportunities for a huge variety of applications, most of which remain probably to be invented. However, RFID is not an objective in itself. It is an enabler for improved efficiencies and conveniences in businesses and in people's life. A lot of research and standardisation efforts have focused on RFID in recent years. Research and standardisation need to continue in order that RFID becomes a commodity seamlessly integrated into an ever growing number of products and services.

References

1. M. Roberti, The History of RFID Technology, RFID Journal (Dec, 2005), http://www.rfidjournal.com/article/view/1338
2. European passive RFID Market sizing. LogicaCMG & GS1, BRIDGE project (2007–2022)

3. P. Chartier, G. van den Akker (Jan, 2010), D1.5 RFID Standardisation State of the art report—Version 3, available on http://www.grifs-project.eu/index.php/downloads/en/
4. A. de Panizza, S. Lindmark, P. Rotter (2010), RFID: Prospects for Europe, item-level tagging and public transportation, EU Joint Research Center, JRC58486 (2010), ISBN 978-92-79-16026-4, http://ftp.jrc.es/EURdoc/JRC58486.pdf
5. ETSI TR 102 649-2 V1.1.1 Electromagnetic compatibility and Radio spectrum Matters (ERM); Technical characteristics of Short Range Devices (SRD) and RFID in the UHF Band; System Reference Document for Radio Frequency Identification (RFID) and SRD equipment; Part 2: additional spectrum requirements for UHF RFID, non-specific SRDs and specific SRDs (2008, 2009)

The World's Smallest RFID Chip Technology

Mitsuo Usami

1 Introduction

The development of effective technology for linking objects and network information systems is being promoted in various fields. Among this technology is semiconductor radio frequency identification (RFID) [1, 2], which is making a significant contribution to promoting a revolution in IT. Product manufacture and distribution require high-quality, accurate inventory control. A small inexpensive chip with RFID capability, called the μ-chip, can be attached to paper media and other small objects to help prevent counterfeit attempts and track products in a market environment. A 0.4 mm × 0.4 mm RFID μ-chip [3], as shown in Fig. 1, which was successfully used for the 2005 World Exposition in Aichi, Japan is available as an example. It features 2.45 GHz operation and a highly reliable 128-bit ROM ID without duplication during the fabrication of 22 million admission-cards. IDs have been developed for use in a reliable authentication system using network based secure ID management. RFID chips have been getting progressively smaller for years and in 2001 Hitachi announced the development of a groundbreaking very small chip. Hitachi has now developed an even smaller chip, which is less than 0.1 mm^2 in size. Recent complementary metal oxide semiconductor (CMOS) fine process technology can produce chips of this very small size, but in our study we had to consider circuit simplification and optimization. Another outstanding feature of the ultra-small RFID chip is an embedded antenna on a chip as shown in Fig. 2.

M. Usami (✉)
Central Research Laboratory, Hitachi Ltd., 1-280 Higashi-Koigakubo, Kokubunji-shi, 185-8601 Tokyo, Japan
e-mail: mitsuo.usami.fc@hitachi.com

H. Casier et al. (eds.), *Analog Circuit Design,*
DOI 10.1007/978-94-007-0391-9_14, © Springer Science+Business Media B.V. 2011

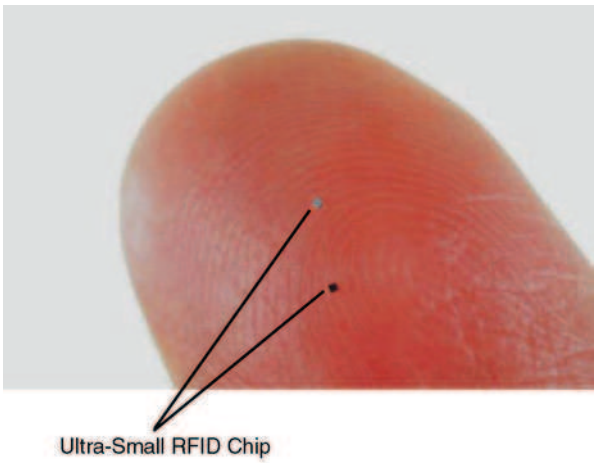


Fig. 1 Typical samples of recently developed ultra-small RFID chips

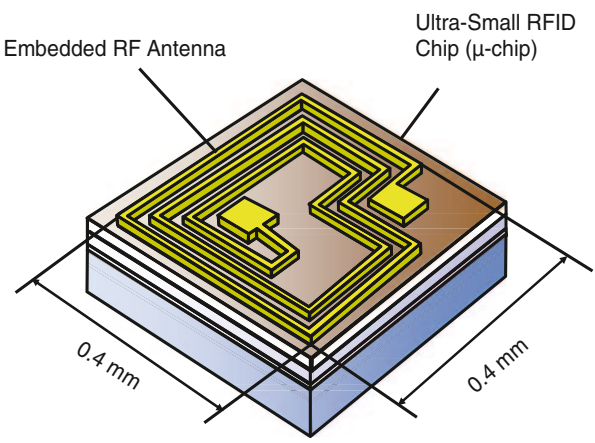


Fig. 2 Embedded antenna structure of ultra-small RFID chip. The small antenna uses 2.45 GHz-band communication without dedicated capacitance

2 Chip Design Concept

2.1 Chip Architecture

The essential elements of μ-chip design are based on the minimum IC chip size for Internet-oriented applications. In practice, μ-chip readers connected to the Internet can easily access world-wide web databases and send timely data to these databases. Therefore, adopting a read-only memory (ROM) technique was considered adequate for implementing an identification number in the μ-chip, which resulted in eliminating writable memory cells and their respective write-control circuits. The

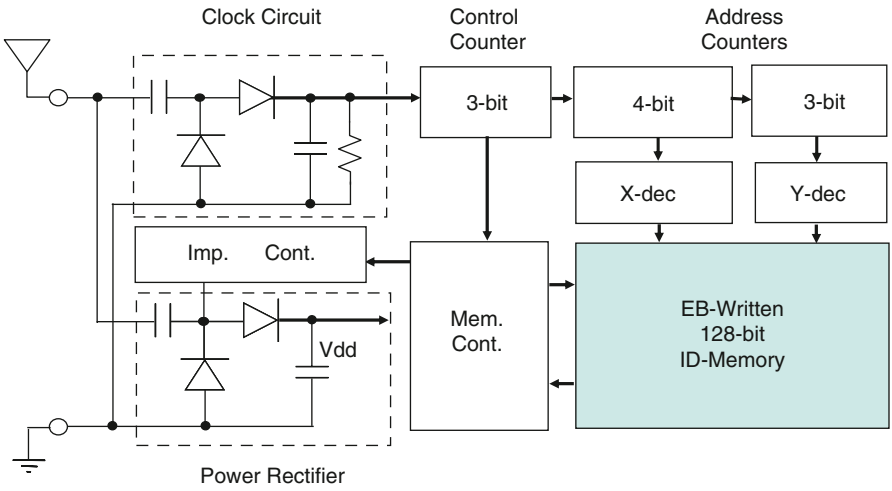


Fig. 3 Circuit structure of ultra-small RFID chip. All of these circuit blocks are made using fully integrated CMOS analog and digital circuit technology

basic structure of the μ-chip is shown in Fig. 3. The ROM size is 128-bit, which is the same length as an IPv6 address. The design policy did not support the anti-collision function in the first μ-chip version. This conceptualistic circuit elimination was crucial to realizing the ultra-small size of the μ-chip.

2.2 Mechanical Endurance

Figure 4 shows that microminiaturization is required to make a reliable IC chip. To enable an IC chip to be used without having to consider the characteristics of the material it is mounted on, such as paper, the IC chip must have adequate stress-resistance. Among the many methods used to measure stress-resistance, the mechanical impact strength test is one of the most rigorous. It is clear that the mechanical strength is higher when the IC chip is smaller than 0.5×0.5 mm^2. Thus, microminiaturization of the μ-chip enabled us to develop an economical and reliable radio-recognition device.

2.3 High-Reliability Electron Beam Written ID Memory

Figure 5 shows circuit details of the 128-bit EB-written memory. Each memory cell consists of one transistor, which has an EB programming connection terminal. The connection information is detected by a simple pre-charge and discharge circuit mechanism. All the drains of the NMOS transistor in the memory cell are commonly connected and store pre-charge electron through a PMOS transistor, which is

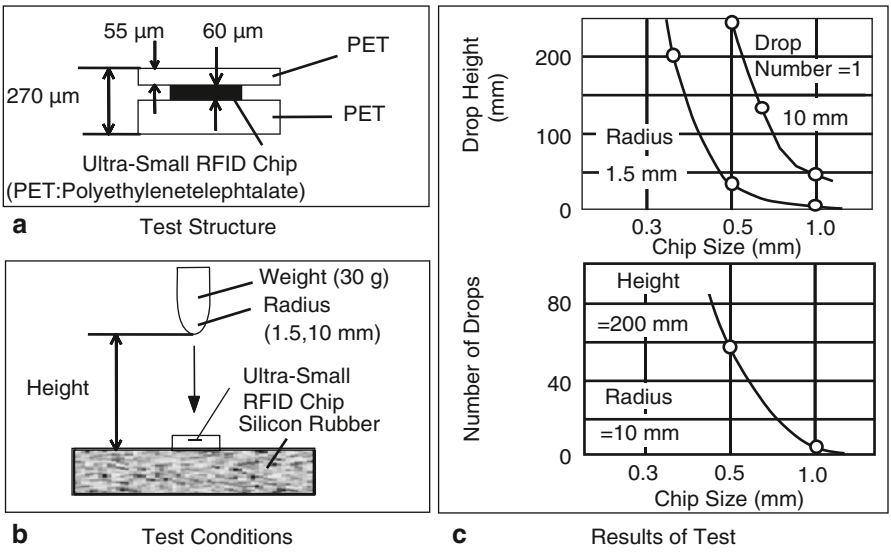


Fig. 4 Mechanical endurance of ultra-small RFID chip. For an IC chip to be usable on soft media such as paper, it must resist stress and remain undamaged during use

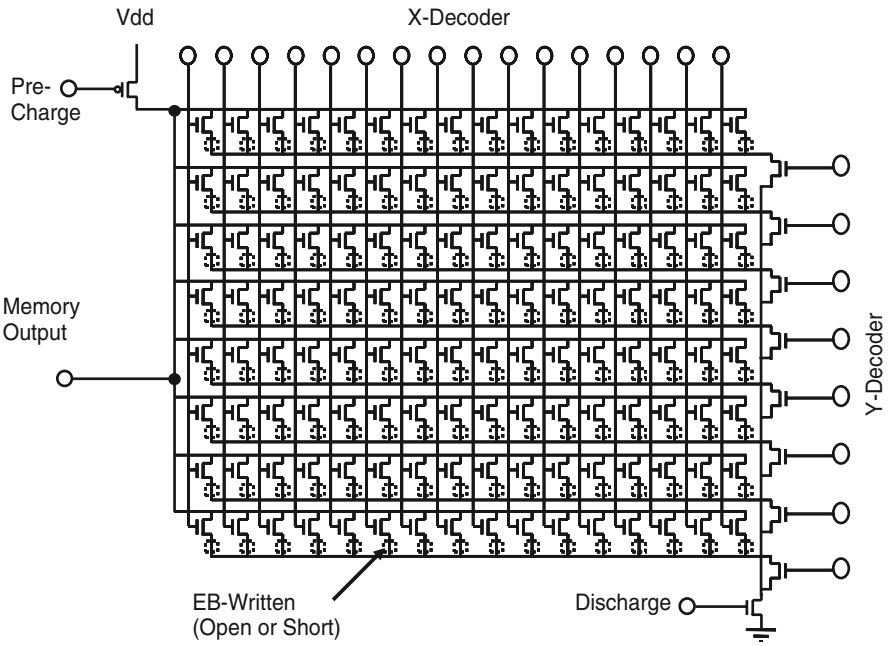


Fig. 5 EB-written high-reliability and ultra-small memory circuit for ultra-small RFID chip

controlled by the memory control circuit. EB programming is performed by open or short connection conditions between each NMOS transistor's source and a common line that is connected to a Y-decoder selection transistor.

This EB-written memory structure has several advantages over other memory technologies. First of all, it can be adapted for use in small RFID devices that use finer device processes because the memory circuit consists of ultra-small transistors and requires no high voltage endurance to write to memory cells. Second, the reliability of RFID devices under various circumstances, such as when exposed to severe temperatures, i.e., 300°C, mechanical stress or radiation, is excellent. Third, maintaining highly secure manufacturing and efficiently controlling the issuing of ID numbers to keep numbering unique during high-volume production is possible because batch EB writing of ID numbers is done during the wafer fabrication process. Finally, EB technology can generate unique ID patterns without using expensive one-time use glass masks.

When using EB writing to ultra-small RFID chips, it is necessary to consider the fabrication throughput. Figure 6 shows chip-grouped EB writing. Chip-grouped EB writing can drastically increase fabrication efficiency. The chip-group size can be up to 10,000 chips per one EB shot. EB write-time is reduced effectively using volume group writing by a factor of 50 compared with indi-

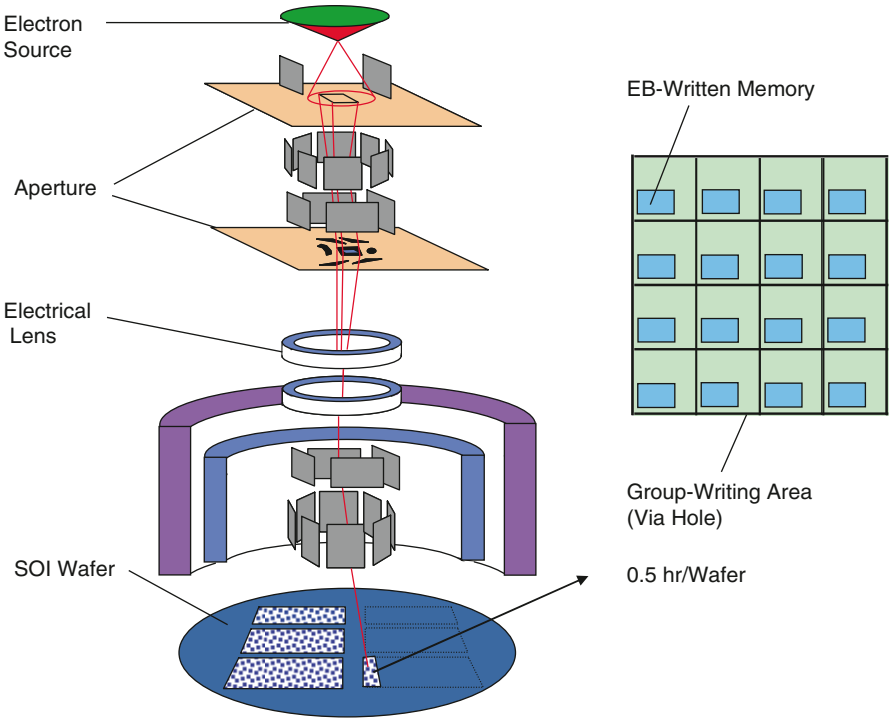


Fig. 6 Group-writing method to realize high-efficiency EB memory writing

vidual writing. The ID number of each chip is controlled by highly-secure sever systems. ID number patterns are merged with the peripheral circuit patterns of the RFID chip and transferred to the EB equipment. Then, group EB writing direct to each chip is done very smoothly as if the writing was being done to a single large chip.

3 Antenna Design Concept

3.1 Embedded Antenna

Embedding an antenna on an RFID chip is an ideal method of reducing the size and cost of RFID devices. The shape of the RFID embedded antenna is similar to a coil pattern because the chip-received electromagnetic energy of induction is greater than the radiation near a reader antenna. The antenna can be designed to cover a relatively small area by using a microwave-band frequency. Figure 2 shows a small 0.4×0.4 mm antenna on a μ-chip. Because of the application of the parasitic capacitance of the antenna, there is no need to design additional resonance capacitance. The antenna is made of gold, using plating on a semiconductor wafer. The design for the internal circuit of an RFID chip with an embedded antenna is exactly the same as that for one with an off-chip external antenna. The maximum communication distance is about 1.2 mm. However, a stable, horizontal communication area with a 2.3-millimeter diameter can be achieved, which is considered to be the practical usage level for close-coupled RFID devices.

3.2 External Antenna on Double-Surface Electrode Chip

Since RFID devices generally consist of chips and an external antenna to extend communication range, the total cost of the identification devices will increase if the cost of assembling an external antenna becomes higher due to the development of ultra small chips. The use of a double-surface electrode [4, 5] on the ultra-thin RFID chip has eased the difficulty of handling small chips and has reduced the cost of attaching RFID antennas. There are two technical problems that occur when a small chip using the conventional approach of a small single-surface electrode is fabricated. One of them is how to precisely position a chip with small surface bumps onto an external antenna's metal terminals, and the other is fabricating narrow space precise antenna patterns. Small double-surface chips can resolve these problems. Because of their double-surface, the surface connection electrodes are connected to the antenna material using a simple con-

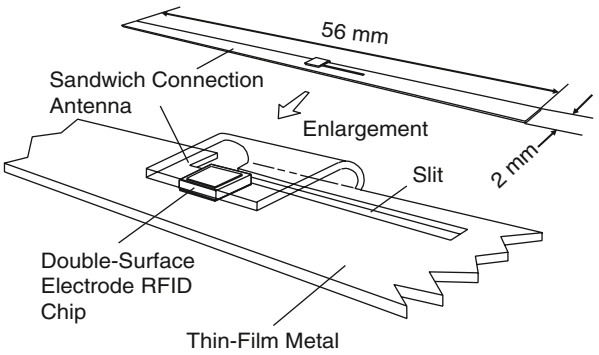


Fig. 7 Antenna connection technology of ultra-small RFID chip. The double-surface electrode RFID chip can be easily connected to the antenna without precise positioning

nection structure. Moreover, the chips do not require horizontal direction regularity for mounting and can be mounted by upside-down free positioning as shown in Fig. 7. So, there are no concerns about positioning. This makes it possible to simultaneously handle a large volume of chips. Another advantage of having a double-surface connection is that each connection area can be designed to be as large as the chip surface in order to reduce connection resistance and enhance connection reliability.

As shown in Fig. 8, the circuit is a double-voltage rectifier circuit based on a capacitor and diode combination, which is similar to the conventional back-bias circuit of dynamic random access memory large scale integration (LSI). The capacitors were designed by using positive channel metal oxide semiconductor (PMOS) device structures. The chip's connection structure is simplified because only one connection electrode port can be placed on each chip surface. This structure is derived from the principle that an RFID chip needs at least two ports to connect with an antenna's terminals. This simple circuit and device structure makes it possible to achieve smaller chip size.

The chip's input impedance is relatively low at about 60 Ω, this being the level required to achieve impedance matching between the chip and the antenna by tun-

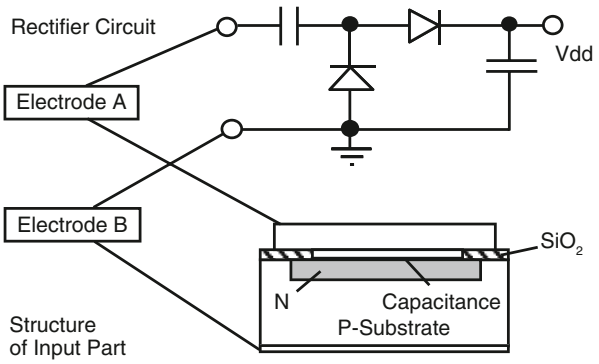


Fig. 8 Device technology of ultra-small RFID chip. The substrate electrode of the chip is able to be connected to one of the terminals of an external antenna

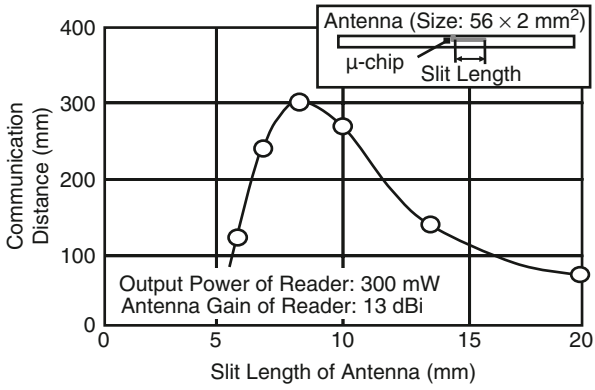


Fig. 9 Communication characteristics of ultra-small RFID chip. The communication distance depends on the slit length of the antenna

ing the length of the impedance-matching slit. Figure 9 shows the communication distance's dependence on the slit length. At a length of 8 mm, a maximum communication distance of 300 mm is achieved giving a reader-output power of 300 mW and a reader-antenna gain of 13 dBi. The simple, straight external antenna attached to the μ-chip is small and thin. The ultra-thin μ-chip (60-μm thick) is connected to this thin antenna using an anisotropic conductive film (ACF) [6]. Using ACF results in a thinner transponder (0.15 mm) than can be produced using conventional wire-bonding technology. The optimized length of the external antenna is one-half of the microwave wave length of 56 mm.

Figure 10 shows the batch-mounting method used to connect the small chip to the antenna. Many chips, in this case 16 chips in a 4 × 4 matrix, can be placed on a

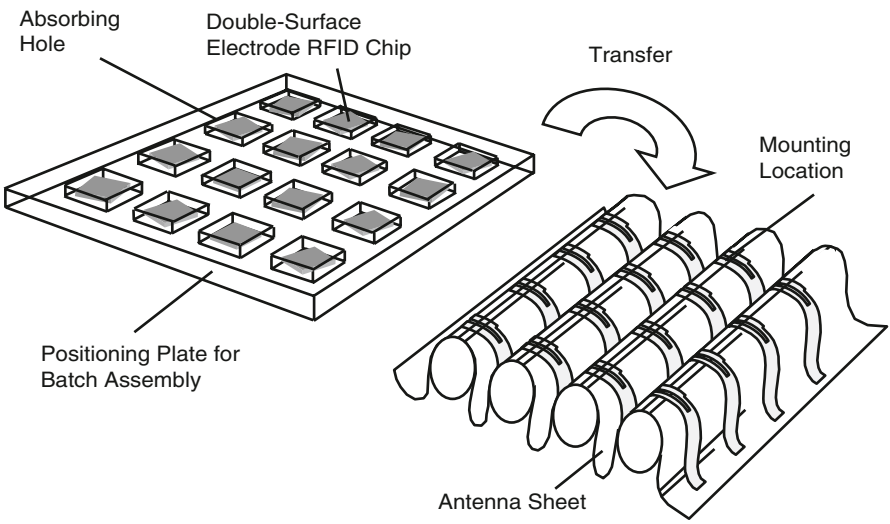


Fig. 10 Assembling technology of ultra small RFID chip. The double-electrode chips are suitable for batch assembly, thus reducing the cost of RFID devices

rough positioning plate. Each chip is dropped slightly into each absorbing hole using vibration, and fixed using a vacuum method. The positioning plate rotates and simultaneously mounts each chip on the antenna sheet, which is folded to match a placing pitch for each chip on the plate. This flexible mounting system makes it possible to mount several chips on to several antenna patterns designed on a substrate film. The size of the matrix can also be extended easily, for example, to 50×50 or more.

4 Silicon on Insulator (SOI) Ultra-thin RFID Chip

Figure 11 compares the front-end structure of the SOI CMOS [7] device with that of a conventional device. The input signal level may reach 2 V or higher when the RFID chip is close to the reader antenna. Therefore, the conventional device needs a two-way or three-way guard ring structure to prevent latch-up from occurring between the well and the substrate. In contrast, in the SOI device, the capacitor and diode devices are insulated by oxide, which enables the RF devices of the chip to be tightly laid out. In the conventional capacitor, the input from the antenna can not be directly connected to the N-well of the device in order to prevent latch-up, and hence the varactor device structure is impractical. However, an SOI CMOS device can use a capacitor structure that is ideal for making a small device. Moreover, the SOI parasitic capacitance between the N-well of the capacitor and the substrate is very small, which resulted in that the dynamic leakage current can be reduced. Lowering the forward voltage drop of the diode is an effective way to increase the efficiency of the rectifier circuit. Therefore, the active body contact technology

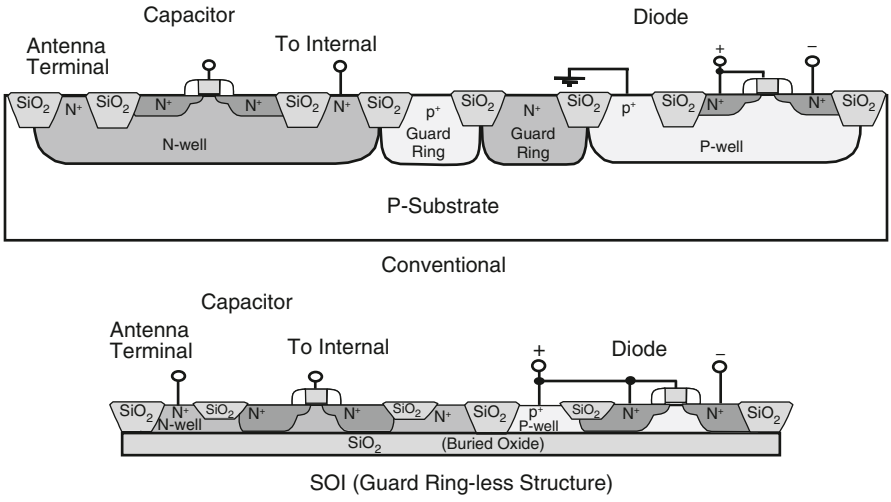


Fig. 11 Comparison of RF front-end device structures

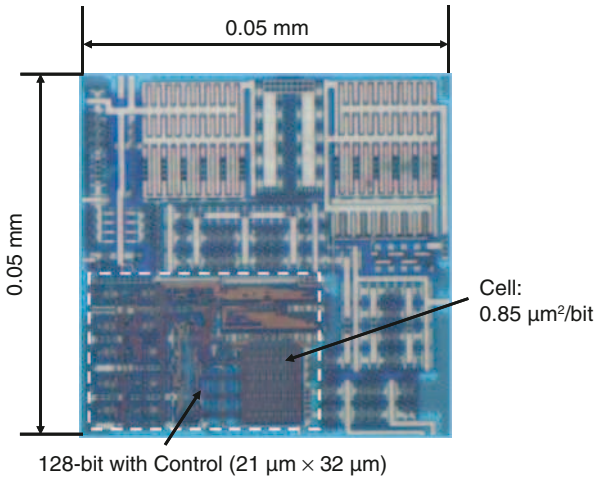


Fig. 12 Microphotograph of ultra-small RFID chip. The top and bottom surfaces are metallized using gold to connect to an external antenna

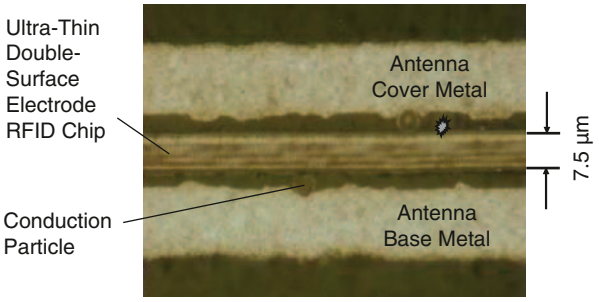


Fig. 13 Partial cross section of antenna connection to ultra-thin double-surface electrode chip

of SOI CMOS can reduce the forward voltage without special devices such as a Schottky diode.

Figure 12 is a microphotograph of the developed chip with the surface electrode pattern removed. This chip was designed with three metallization levels.

Figure 13 shows a part of a cross section of the antenna connection to the chip. A thin ACF connection technique was also used to fabricate the RFID inlet, which is thinner than 50 μm, for versatile individual recognition applications to thin media such as paper. The inlet holds the intermediate fabrication unit of the device, which consists of the chip and its antenna [8].

5 Conclusions

Ultra-small RFID chips are expected to be used in future volume-product applications such as bar-code replacement. It is not easy to realize the instant world-wide extension of RFID. However, semiconductor and wireless circuit technology progress will someday overcome the function and cost problems of RFID chips.

Acknowledgments The author would like to express special appreciation to S. Asai, R. Imura, and A. Isobe of Hitachi for encouraging this research, and to thank Y. Tsunemi, K. Takaragi, C. Okamoto, A. Nakano, H. Hayakawa, F. Murai, H. Takase, I. Sakama, T. Ookuma, O. Horiuchi, T. Sasaki, K. Ootsuka, T. Ishige, H. Ishizaka, K. Tasaki, H. Matsuzaka, N. Matsuzaka, T. Satou, Y. Hara, M. Ashizawa, H. Tanabe, A. Sato, T. Iwamatsu, and S. Maegawa for their significant contributions to this research and the development work.

References

1. U. Kaiser et al., A low-power transponder ic for high-performance identification systems. IEEE J. Solid-State Circuits. **30**(3), 306–310 (Mar 1995)
2. D. Friedman et al., A low-power cmos integrated circuit for field-powered radio frequency identification tags. ISSCC Digest of Technical Papers, Feb 1997, pp. 294, 295
3. K. Takaragi et al., An ultra small individual recognition security chip. IEEE micro. **21**(6), 43–49 (Nov/Dec 2001)
4. M. Usami et al., Powder LSI: an ultra small rf identification chip for individual recognition applications. ISSCC Digest of Technical Papers, Feb 2003, pp. 398, 399
5. S. Briole et al., AC-Only RFID tags for barcode replacement. ISSCC Digest of Technical Papers, Feb 2004, pp. 438–439
6. M. Usami, Thin silicon chips and ACF connection technology for contactless IC cards. Proceedings of IMAPS 32nd International Symposium on Microelectronics, Oct 1999, pp. 309–312
7. Y. Hirano et al., Impact of 0.10 μm SOI CMOS with body-tied hybrid trench isolation structure to break through the scaling crisis of silicon technology. IEDM Technical Digest, Dec 2000, pp. 467–470
8. M. Usami et al., An SOI-Based 7.5 μm-Thick 0.15×0.15 mm^2 RFID Chip. ISSCC Digest of Technical Papers, Feb 2006, pp. 308, 309

RF and Low Power Analog Design for RFID

Raymond Barnett

1 Introduction

The RFID system, shown in Fig. 1, has a master reader to communicate to and gather information from one or more RFID tags.

The tag is attached to an item to be tracked and is used to identify the item in a number of different applications [1]. The reader and tag communication is half duplex, controlled by the master reader. RFID tags can be active or passive. The active tag requires a battery to operate, whereas a passive tag receives both power and data through the RF link. Two typical frequency bands used are HF, or 13.56 MHz magnetic field powered, and UHF or electromagnetic field powered, operating around 900 MHz. Some RFID applications are extremely cost sensitive; For example the RFID tag intended for bar code replacement is targeted for less than $ 0.05. Passive HF tags are typically operated up to tens of centimeters, whereas UHF tags have pushed ranges to 10 m and beyond. The passive tag IC requires several sub-blocks [2–6] as illustrated in Fig. 2.

The Rectifier is used to convert the incoming RF field to a DC power supply to operate the tag processing circuits. Since the RF field can vary several orders of magnitude, a RF Limiter, DC Limiter and DC Regulator are required to produce a stable DC power supply, V_{DD}. The tag demodulates and decodes an incoming data command typically in Amplitude Shift Keying (ASK) format. The tag must interpret the data and send back a appropriate response, usually in the form of a ID number stored in Non-Volatile Memory (NVM). As the reader "listens" for responses it transmits a continuous RF power wave with no modulation. The tags respond by using intelligent reflection of the continuous power signal using an antenna modulator. In the literature this is referred to as backscatter transmission, load modulation, antenna mismatching or impedance modulation. Additional support circuits on the IC are clock generator, power on reset (POR), bias, non-volatile memory and digital back end state machine. This paper addresses some of the main RF and analog circuits.

R. Barnett (✉)
Texas Instruments, Texas, USA

H. Casier et al. (eds.), *Analog Circuit Design,*
DOI 10.1007/978-94-007-0391-9_15, © Springer Science+Business Media B.V. 2011

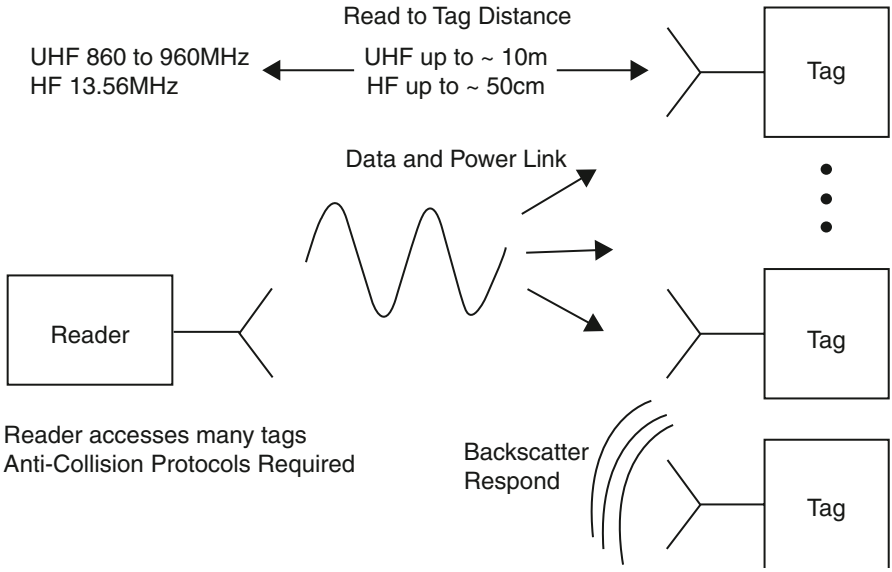


Fig. 1 The RFID system showing the master reader and several tags

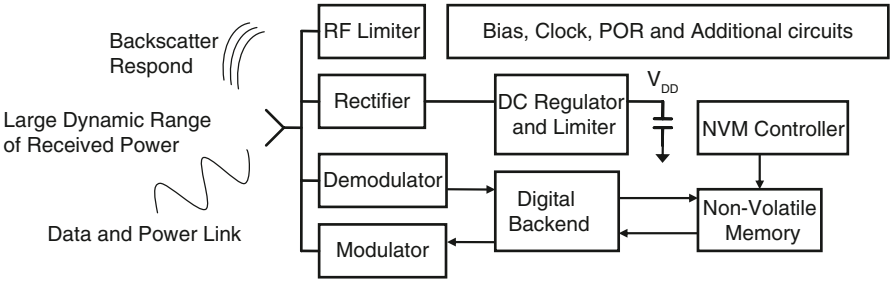


Fig. 2 The RFID tag IC and the various sub-blocks

2 Rectifiers for UHF and HF RFID Tags

The DC power required to operate the transponder IC is generated using a rectifier that converts the incoming RF signal to a DC supply voltage. The rectifier must produce enough DC output voltage to operate the IC, typically around 1 V. At the minimum operating voltage, the rectifier must be as efficient as possible in order to maximize communication distance. In the case of HF RFID, the magnetic near-field is received by the receiving coil antenna and the HF rectifier converts the AC voltage to a DC output voltage. For UHF, the RF electromagnetic far-field is received by the antenna and converted to a DC voltage using a RF rectifier. This section will cover HF and UHF rectifier design.

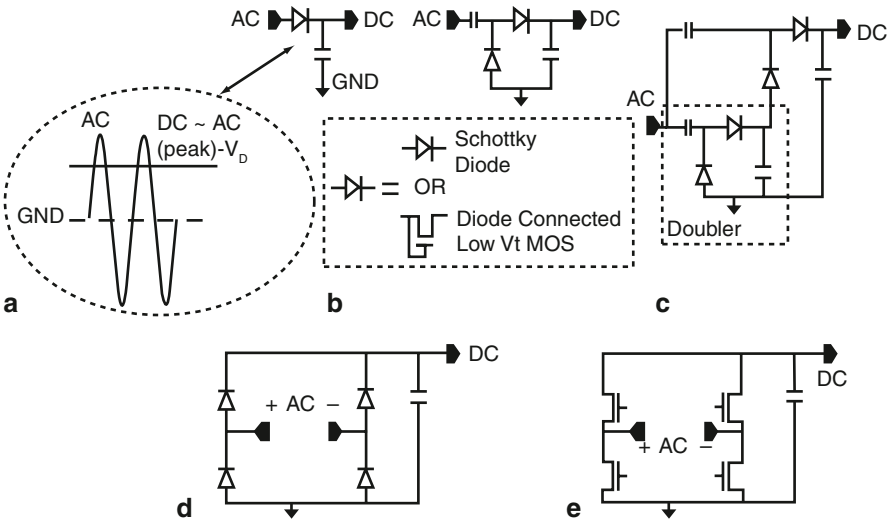


Fig. 3 Rectifier topologies used in HF and UHF RFID tags. **a** Single diode rectifier. **b** Dual diode rectifier "doubler". **c** Multi-stage rectifier cascade of "doublers". **d** Diode bridge rectifier. **e** MOS bridge rectifier many combinations of NMOS and PMOS many combinations of gate drive

2.1 General Rectifier Topologies

Figure 3 shows various Rectifier topologies each labeled "A" through "E".

The most basic rectifier topology, "A", is the diode rectifier comprised of a single diode and capacitor. The antenna is connected to port *AC* and *GND*. With zero load current drawn from the *DC* port, the maximum output voltage on port *DC* is determined by the input *AC* voltage peak swing minus the diode turn on voltage, V_D. This is illustrated by the waveform in Fig. 3. Increasing load current reduces the output voltage from the maximum value. Efficiency is improved by reducing the diode drop, thus low turn on voltage Schottky diodes are the preferred technology for the diode rectifier. Improvements to the diode rectifier are the diode doubler and multi-stage rectifier shown as "B" and "C" respectively. The diode doubler ideally produces twice the DC output voltage as the single diode rectifier. During the negative AC cycle, charge is stored on the input coupling capacitor. On the positive cycle, the stored charge is transferred to the load, and twice the peak *AC* voltage is delivered to the *DC* port. Thus for the diode doubler, the *DC* port voltage is two times the output voltage when compared to the single diode rectifier. The multi-stage rectifier is a cascade of diode doublers to multiply the output voltage even further [3], at the cost of increasing RF loss and DC output impedance. Detailed design and analysis of multi-stage rectifiers can be found in [7, 8], and the next section will summarize the results of [8]. Also to note, when Schottky diodes are not available in the process, the diode used in the rectifiers may be replaced by a low or zero Vth MOS device. Even a well designed Schottky diode will have 100–200 mV forward voltage with

typical RFID received power and loads. MOS devices attempt to mimic "zero turn on" voltage devices, or "ideal diodes" with low forward drop [5] and [9].

Rectifiers shown in "D" and "E" are typically used for HF applications where voltage multiplication is not required. "D" is a diode bridge full wave rectifier. The DC output is similar to the single diode rectifier shown in "A", except both positive and negative peaks transfer charge to the load capacitor. Thus the cycle to cycle charge loss on the DC load capacitor due to load current is less than the single diode rectifier. Another way of stating this is the equivalent DC Thevenin output impedance is ½ the value of the single diode rectifier. Finally "E" shows the MOS version of the full wave rectifier. The gate connections are not shown as there are many variations. A few of the variations will be discussed in Sect. 2.3. The basic idea for using MOS devices is to control the gate voltages such that the MOS devices mimic ideal diodes. This is done by sensing the forward current condition and turning on the MOS device to conduct in the forward direction. As the current reverses, the MOS device is turned off. The advantage of MOS devices over diodes is the MOS can be controlled to operate like a diode with lower forward voltage drop. This results in higher output voltage and hence better efficiency. Since the MOS gates are controlled at defined time instances, this is often referred to as a synchronous rectifier. The challenge in implementing the synchronous rectifier is determining the appropriate time to turn on the MOS devices so as to mimic ideal diodes. Some methods to do this will also be discussed in Sect. 2.3. The next two sections will discuss UHF and HF rectifiers respectively.

2.2 UHF Rectifier Design

The power transmitted by the reader in UHF RFID is limited by regulations, and hence the received voltages are a few hundred mV for read ranges of 10 m. Therefore, if maximum read range is desired, UHF designs require a multi-stage rectifier to multiply the few hundred mV RF signal up to a DC supply of 0.8–1 V. This section will briefly review the multi-stage rectifier analysis given in [8], which expands the original Dickson charge pump analysis [10]. The resulting equation solves for the rectifier DC output, V_{DC}, as a function the available power from the antenna, antenna radiation resistance, number of diodes in the rectifier, coupling capacitor value, parasitic capacitive loss terms, parasitic resistive loss terms, diode physical properties and load current. This single equation can be used to quickly analyze the multi-stage rectifier circuit when connected to an antenna via a matching circuit. The analysis method reduces a non-linear behaving circuit with RF and DC signals to an equation for the DC output voltage that is a sum of gain and loss terms involving all the major design parameters of the rectifier. The DC output voltage resulting from the equation matches non-linear time domain simulations to within 5–10%. The equation also gives intuitive insight into the various tradeoffs in the design parameters. One major insight is increasing the antenna radiation resistance indefinitely does not lead to continually higher rectifier DC output voltage. Another

major insight is increasing the number of diodes in the rectifier indefinitely does not lead to continually increasing output voltage. A review of the Dickson equation is next, followed by a review of the expanded analysis in [8].

The Dickson analysis is the most widely quoted equation in RFID publications because it reveals an intuitive understanding of the effects of the number of diodes in the rectifier, the frequency of operation, the input AC voltage swing, and the load current variations. This equation was originally developed for low frequency operation where the AC input voltage is large and the diode turn on voltage can be considered as a fixed voltage. However, it is not a good predictor of RF to DC conversion at UHF frequencies because of low input voltage swings, antenna impedance effects, diode drop variation with load current, and component losses. The work in [8] addresses these issues. It does this by accounting for the non-linear diode drop in an intuitive fashion, determining the fundamental mode input impedance of the non-linear rectifier under various parameters such as load current and number of diodes, accounting for significant RF loss components, and then solving for the available RF power to DC output voltage transfer function equation. The final solution requires three analysis steps before solving for the final equation. The three steps will be summarized next and the final equation given. Readers interested in the full derivation should consult [8].

The first step in the analysis is to account for the diode non-linear voltage drop as a function of the DC load current. Then use this in place of the fixed diode drop term in the Dickson equation. The results of this step are summarized and illustrated in Fig. 4.

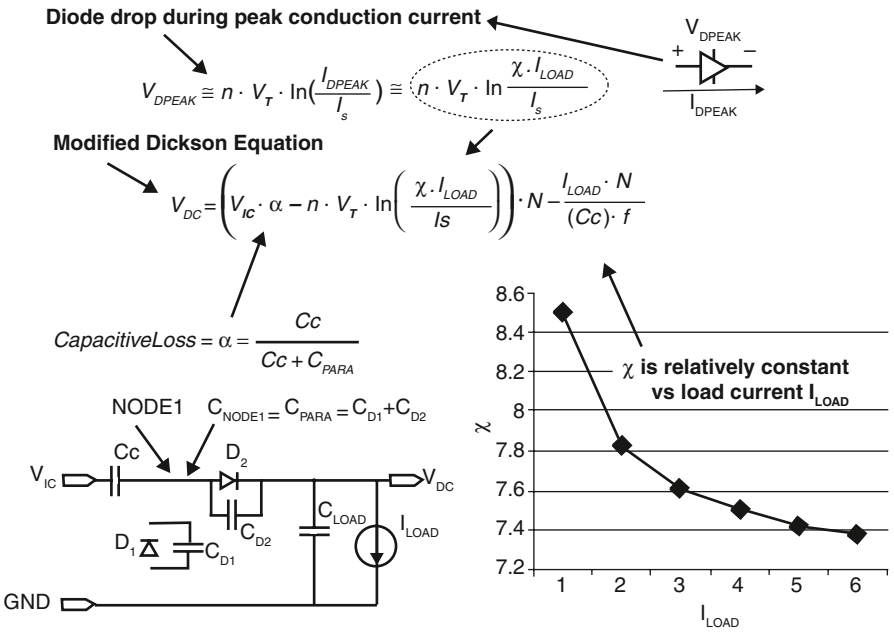


Fig. 4 Accounting for the non-linear diode drop in the Dickson equation

The top of Fig. 4 shows two equations, the equation for the peak diode forward drop, V_{DPEAK}, and the modified Dickson equation showing the replacement of V_{DPEAK} for the fixed diode drop term. The peak AC input current, I_{DPEAK}, is approximately equal to an empirical factor, χ, times the DC load current, I_{LOAD}. The empirical factor is shown to be relatively constant over typical RFID load currents at minimum sensitivity, where high efficiency matters most. A graph of load current vs. χ is shown in the figure. Also depicted in Fig. 4 is a diode doubler schematic, the basic building block of the multi-stage rectifier, showing components important in determining the DC output voltage, V_{DC}. In the Dickson equation, V_{DC} is the DC output voltage at steady state, V_{IC} is the zero to peak AC voltage at the rectifier input, Cc is the coupling capacitor, C_{PARA} is the sum of parasitic diode capacitances C_{D1} and C_{D2}, I_S is the diode saturation current, Vt is the thermal voltage equal to kT/q, I_{LOAD} is the load current, f is the carrier frequency and N is the total number of diodes, 2 for the doubler. The Dickson equation is derived using a charge conservation analysis while considering the charge loss per AC cycle due to I_{LOAD}. Since C_{LOAD} is assumed much larger than Cc, C_{LOAD} does not appear in the final AC to DC steady state equation. However, C_{LOAD} does contribute to the time it takes to charge to the steady state value. The modified Dickson result is interpreted as follows. The AC input voltage, V_{IC}, is multiplied by a capacitive loss term, α, and a load variable diode forward voltage is subtracted from the result. The resulting "voltage" is multiplied by the number of diodes in the multi-stage rectifier, N. Finally, a loss term that is proportional to the load current, I_{LOAD}, and the number of diodes, N, is subtracted from the result to yield the final output voltage, V_{DC}. The loss term that is proportional to I_{LOAD} can be viewed as a Thevenin output "resistance" term. The more load current, the larger the voltage drop across the Thevenin "resistance", and the lower the resulting DC output voltage. This term is a direct result of the charge conservation analysis and has the same form as a switched capacitor resistor, with "resistance" inversely proportional to frequency, f, and the "switched capacitor", Cc.

The second step in the analysis is to determine the rectifier input shunt resistance, R_{IC}, and the input shunt capacitance, C_{IC}. These components are used to determine how the open circuit antenna voltage is translated to the rectifier input voltage. Once the rectifier input voltage is known, the modified Dickson equation is used to determine the final DC output voltage. Figure 5 shows a model of a dipole antenna, matching inductor and IC input, along with the important signals used in determining R_{IC}. The antenna is modeled by V_{ANT}, R_{ANT} and C_{ANT} elements, where V_{ANT} is the open circuit voltage of the antenna, R_{ANT} is the antenna radiation resistance, and C_{ANT} is the shunt capacitance of the antenna. V_{ANT} can be expressed as a function of the radiation resistance of the antenna, R_{ANT}, and the available power, P_{avail}, by the following equation

$$V_{ANT} = \sqrt{P_{avail} \cdot R_{ANT} \cdot 8}. \tag{1}$$

For a given power available from the antenna, an increase of R_{ANT} is required to increase V_{ANT}, and it is desirable to increase V_{ANT} to overcome the threshold voltage of

the diodes used in the rectifier. As stated previously, the final outcome of this analysis shows that even though the open circuit antenna voltage continually increases with increasing R_{ANT}, the DC output of the rectifier does not continually increase.

Inductor L_{EXT} is used to resonate out the total capacitance of the antenna, C_{ANT}, and the IC, C_{IC}, at the desired frequency band around 900 MHz. Thus analyzing the circuit of Fig. 5 at the resonant frequency the reactive components cancel and can be removed from the circuit. The transfer function equation relating *Pavail* to the voltage at the IC terminals, V_{IC}, is given by

$$V_{IC} = \sqrt{P_{avail} \cdot R_{ANT} \cdot 8} \cdot \frac{G_{ANT}}{G_{ANT} + G_{IC}}. \tag{2}$$

Note that V_{IC} increases with G_{ANT}, however, maximum power transfer occurs when G_{ANT} is equal to G_{IC}. It is important to remember that when connected to an antenna, impedance matching is important as this maximizes the power transfer from the antenna, even though the voltage V_{IC} is not maximized. This is in contrast to the case if V_{IC} was a low impedance AC voltage source, where the highest V_{IC} would maximize the rectifier DC output voltage. In order to determine R_{IC}, the equivalent fundamental impedance of the diodes at the RF input frequency is derived. The fundamental impedance is of importance since the front end is a tuned circuit and the "harmonic impedances" are filtered out by the tuned network. The waveforms in Fig. 5 show how the higher order components of the input AC current are filtered by the inductor and capacitor at resonance. The full derivation of R_{IC} and C_{IC} is given in [8] but the general concept to derive R_{IC} is as follows.

The rectifier input voltage for the low sensitivity region of operation is quite sinusoidal. Low sensitivity is defined here as the region where no RF or DC limiting action is occurring; approximately when Pavail < 0 dBm. The desire is optimize the rectifier when the power level is −14 to −16 dBm, so a sinusoidal voltage in this region of operation is a valid assumption. The current however is non-linear at point B. This is because on each cycle, the diodes conduct currents during the peak voltage and ceasing conduction shortly after the peak voltage. This is illustrated by the sharp current pulses, also referred to as small conduction angle current pulses, shown in Fig. 5. As long as the conduction angle is small, the actual shape of the pulses is not important. This is key to the analysis, as the shape of the current is quite complex and is not solvable without using Bessel functions and ultimately numerical calculation. The shape is not important, because the current at point A is the fundamental component of the non-linear current at point B, and by using the cosine Fourier transform as outlined in [8], it is simply the area under the curve that is important. The result is the fundamental component of the non-linear current for each diode is approximately equal to twice the DC load current, as illustrated in Fig. 5. An important note is even though the shape of the current pulse at point B is not known, and the output voltage of the rectifier is not known, the fundamental resistance of the diodes at point A can be calculated by simply knowing the load current, voltage swing and the number of diodes.

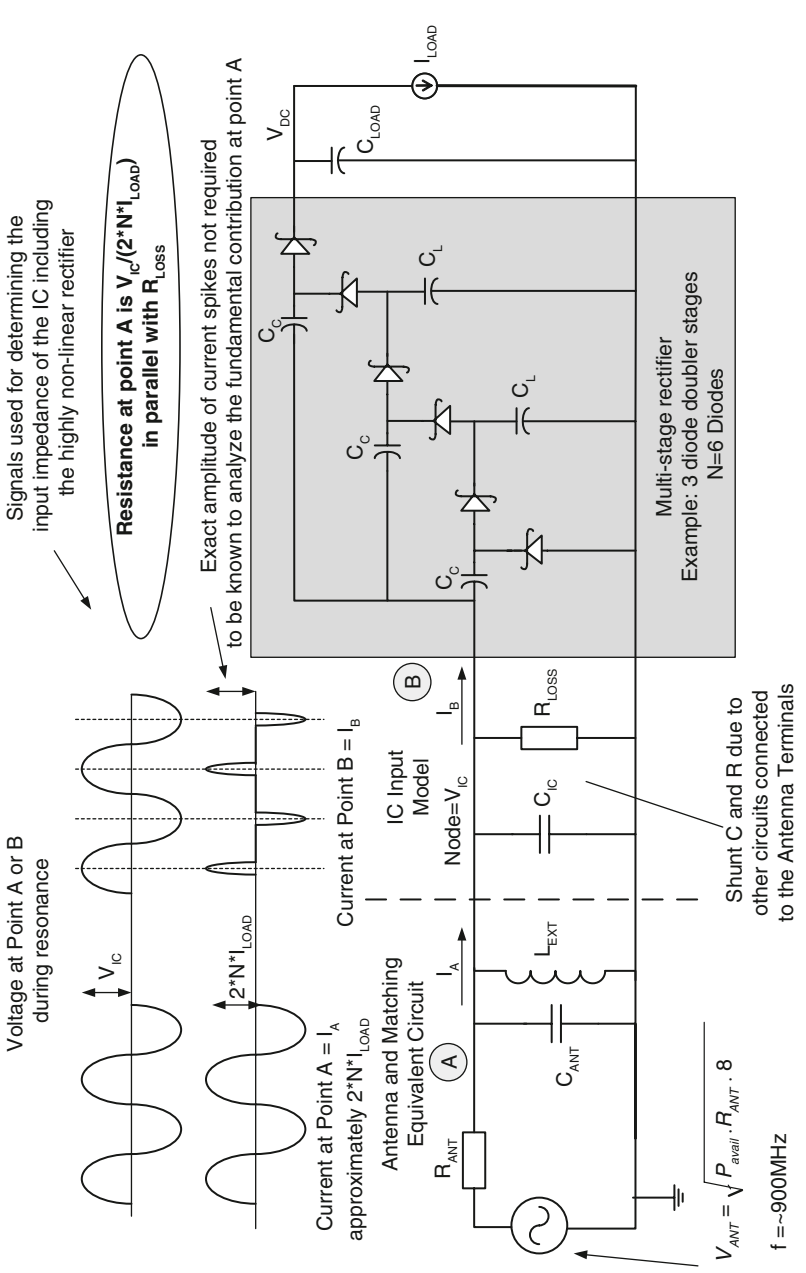

Signals used for determining the input impedance of the IC including the highly non-linear rectifier
Resistance at point A is $V_{IC}/(2*N*I_{LOAD})$ in parallel with R_{LOSS}
Exact amplitude of current spikes not required to be known to analyze the fundamental contribution at point A
Voltage at Point A or B during resonance
V_{IC}
$2*N*I_{LOAD}$
Current at Point A = I_A approximately $2*N*I_{LOAD}$
Current at Point B = I_B
Antenna and Matching Equivalent Circuit
IC Input Model
Node=V_{IC}
R_{ANT}
C_{ANT}
L_{EXT}
I_A
I_B
C_{IC}
R_{LOSS}
C_C
C_L
V_{DC}
C_{LOAD}
I_{LOAD}
Multi-stage rectifier
Example: 3 diode doubler stages
N=6 Diodes
Shunt C and R due to other circuits connected to the Antenna Terminals
$V_{ANT} = \sqrt{P_{avail} \cdot R_{ANT} \cdot 8}$
f =~900MHz

Fig. 5 Simplified model of the antenna and rectifier impedance

The third and final step before solving the full transfer function equation is to include appropriate loss components. The reader is referred to [8] for the details of including the loss components and the derivation of the P_{avail} to V_{DC} transfer equation which is given by Eq. 3.

$$V_{OUT} \cong \underbrace{\sqrt{P_{avail} \cdot R_{ANT} \cdot 8} \cdot \alpha \cdot N}_{\text{A}} - \underbrace{\sqrt{P_{avail} \cdot R_{ANT} \cdot 8} \cdot \alpha \cdot N \cdot \frac{G_{LOSSFIX}}{G_{ANT}}}_{\text{B}} - \underbrace{\sqrt{P_{avail} \cdot R_{ANT} \cdot 8} \cdot \alpha \cdot N^2 \cdot \frac{G_{LOSSN}}{G_{ANT}}}_{\text{C}}$$
$$\underbrace{-2N^2 \cdot (I_{LOAD} \cdot R_{ANT})}_{\text{D}} + \underbrace{2N^2 \cdot (I_{LOAD} \cdot R_{ANT}) \cdot \frac{G_{LOSSFIX}}{G_{ANT}}}_{\text{E}} + \underbrace{2N^3 \cdot (I_{LOAD} \cdot R_{ANT}) \cdot \frac{G_{LOSSN}}{G_{ANT}}}_{\text{F}}$$
$$\underbrace{-n \cdot V_T \cdot \ln\left(\frac{\chi \cdot I_{LOAD}}{Is}\right) \cdot N}_{\text{G}} - \underbrace{\frac{I_{LOAD} \cdot N}{(C_C) \cdot f}}_{\text{H}} \quad (3)$$

The interpretation of each term of the equation is as follows. Term "A" represents the maximum gain of the rectifier. This is given by the input voltage, which is equal to $\sqrt{P_{avail} \cdot R_{ANT} \cdot 8}$, times the number of diodes, N, while including the capacitive loss term, α, typically 0.8–1. Term "H" is the output impedance as given by the Dickson analysis and can usually be ignored in the design of rectifiers at UHF, as the other output loss terms are more significant. Term "G" is a result of including the nonlinear forward voltage drop of the diode. Terms "B" and "C" are additional loss terms that are independent of the load current, I_{LOAD}. These are input voltage divider losses when internal rectifier losses are transformed to the input. Finally, terms "D", "E" and "F" depend on the output load current and therefore can be viewed as additional nonlinear output impedance terms. Term "D" is a significant loss term, whereas terms "E" and "F" actually add to the output voltage. However, "E" and "F" are relatively insignificant for typical loss values. Note these terms contain N^2 and N^3 factors and are a result of the analysis of the input impedance of the rectifier. Effectively, terms "D", "E" and "F" connect the output load current to the source resistance, R_{ANT}. Thus an output voltage drop or loss term proportional to $I_{LOAD} * R_{ANT}$ is present, as if the DC load current at the output of the rectifier were flowing through the antenna radiation resistance causing a DC voltage loss at the output of the rectifier.

As stated previously, first order analysis of the multi-stage rectifier may lead one to conclude that increasing both N and R_{ANT} lead to higher DC output voltage, however the result of (3) is increasing N does not always increase the output voltage because of the higher order loss terms involving N. The second result is that increasing R_{ANT} does not always increase the output voltage, even when no RF losses are considered. This is seen in the loss term "D", which is directly proportional to R_{ANT}, whereas the gain term "A" is proportional to the square root of R_{ANT}. In addition to the insights given by (3), the equation can be differentiated with respect to N and solved for a given process to find the optimal number of diodes needed to maximize

the output voltage. Additionally (3) can be used to find the efficiency as outlined in [8]. In summary, it should be noted that efficiency of the rectifier is meaningful in RFID, only when combined with the minimum output voltage required by the rectifier load. This minimum output voltage is the supply voltage required to operate the RFID processing circuitry. Next, HF rectifier topologies will be discussed.

2.3 HF Rectifier Design

HF RFID has typically targeted a maximum read range of tens of cm, and therefore, the minimum received voltage levels are normally several volts. As a result, multiplying the input voltage is not required. Thus, HF designs can operate with single or dual diode rectifiers and produce ample voltage to operate the IC circuits. The single diode rectifier of Fig. 3a or the diode bridge rectifier shown in Fig. 3d are popular choices for HF rectifier designs. The bridge rectifier transfer function is similar to the single diode rectifier, however, the DC output impedance is lower and the output voltage has lower ripple voltage. This is due to full-wave rectification as opposed to half-wave rectification achieved by the single diode rectifier. In many CMOS processes a more efficient rectifier may result from using MOS transistors as synchronous switches in place of diodes. For example, the MOS version shown in Fig. 3e is used in order to overcome the diode forward voltage losses. There are many combinations of PMOS/NMOS and gate drive for the MOS version. The gate drive can be diode connected, and work the same as the diode bridge, or some of the gates can be controlled such that the MOS devices operate as switches, with low voltage drop, thus improving efficiency. Figure 6 shows two examples of MOS bridge rectifier architectures.

Figure 6a shows a "diode connected" MOS bridge rectifier. The structure is configured the same as the diode bridge rectifier of Fig. 3d but using MOS devices in place of the diodes. The MOS devices contain parasitic diodes in the same directions as the standard diode rectifier. The MOS devices are "diode connected" with gate tied back to drain or source (note in symmetric MOS devices the drain and

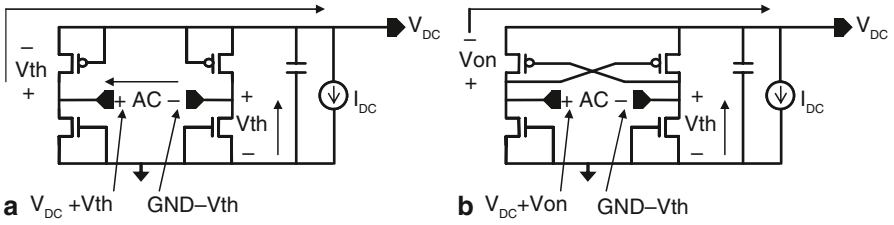


Fig. 6 Examples of MOS based HF rectifiers. **a** Diode connected MOS bridge body diodes in parallel with MOS diodes body diodes not shown. **b** Diode connected NFET cross coupled PFET

source terminology are interchangeable), and thus the "diode connected" MOS is in parallel with the parasitic diode. If low Vth MOS devices are used, the MOS devices conduct current before the parasitic diodes. Thus in Fig. 6 Vth is shown as the drop across the device instead of the parasitic diode drop, V_D. The voltage between the ports labeled "AC" is the coil voltage. The coil voltage is the received signal that is dependent on reader transmitted power, distance, coil geometries, etc. The peak coil voltage is converted to the DC output voltage V_{DC}. In steady state with no load current the DC output is approximately V_{AC}—2 Vth, with Vth dependent on MOS technology and typically ranging from 0.3 to 0.7 V. Output voltage and efficiency can be improved by using the lowest Vth possible. The circuit shown in Fig. 6b replaces the "diode connected" PMOS devices with a cross coupled connection. The PMOS are operated as switches and thus the voltage drop across the devices while conducting is lower than Vth, thus improving efficiency. The voltage drop is labeled Von, and this equal to the on resistance of the FET, Ron, times the current flowing through the device. The device Ron is a function of device size, gate drive level and MOS technology parameters. With typical RFID tag parameters Von is less than 100 mV. Figure 6 shows two examples of MOS based designs and other useful structures exist, such as cross coupled PMOS and NMOS, and other combinations. The rectifiers shown in this section are useful when the application does not require voltage multiplication. Voltage multiplication is not required in applications if the received peak voltage is greater than the supply voltage required by the IC. These applications include LF (~100 kHz) designs, HF, and short range UHF. The function of the rectifier is to convert the incoming RF peak voltage to a DC supply voltage. Since received power levels can vary several orders of magnitude, damaging voltage levels may occur in the IC. Therefore when excess power is available, additional circuits are required to divert RF power and regulate the DC supply generated by the rectifier, in order to protect the IC from over-voltages. The next section presents an overview of Power Management architecture and design for RFID.

3 Power Management Design for RFID

Managing the power levels in an RFID system can be challenging due to the large variation of the incoming RF power from the antenna, combined with the need to use low voltage IC technology. On one hand, at long distances from the reader, the tag receives a weak RF signal and is required to convert the RF power to a DC supply with the highest efficiency possible. On the other hand, the IC may be very close to the reader, and experience high voltages due to the high RF fields. Therefore the field power varies several orders of magnitude. Designs such as the UHF Power Harvester shown in Fig. 7 use a multi-step approach to manage the high RF power levels.

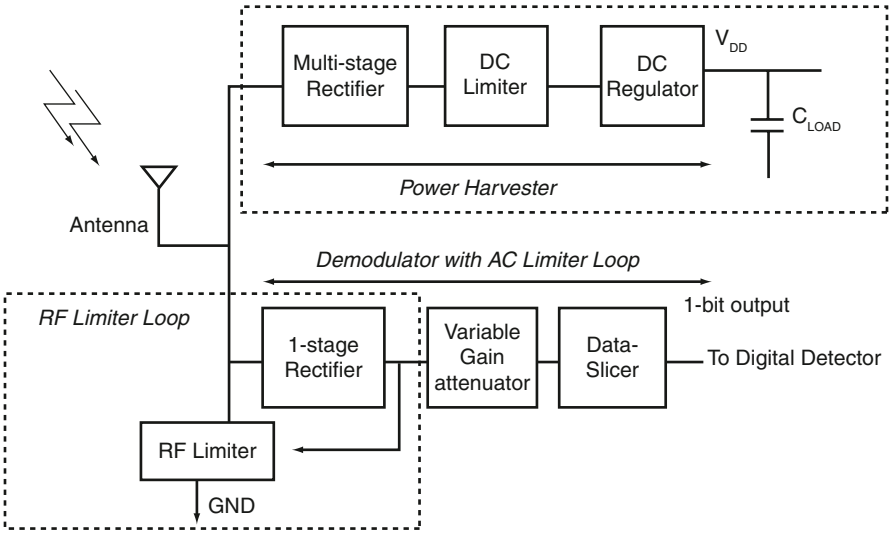


Fig. 7 Power and Data Receiver in RFID showing the Power Harvester Path and RF Limiter Loop

Figure 7 shows the antenna connected to the Power Harvester path and the data receiver, or Demodulator path. The Power Harvester consists of the Multi-stage Rectifier, DC Limiter and DC Regulator producing the V_{DD} supply on an internal capacitor, C_{LOAD}. The RF Limiter is another sub-circuit of the power management block. During weak RF fields, the RF Limiter is inactive, the Crude DC Limiter is inactive, and the output voltage from the Rectifier is transferred to the V_{DD} supply with as little voltage loss as possible. As the field strength increases, it is necessary to limit the voltages and divert power from the IC. The first line of defense against high RF fields is to reflect incoming power from the antenna using the RF Limiter. The Limiter uses a feedback loop to keep the antenna voltage below a certain threshold as required by the process. This is typically in the range of 1.5–3.6 V peak depending on the IC technology used. The second line of defense is the DC Limiter, which reduces the DC output voltage of the rectifier in an unregulated fashion. The final regulation step is to implement fine control of the output voltage to a level acceptable by the IC technology. This function is performed by the DC Regulator. The next sections will outline architectures for the RF Limiter, DC Limiter and DC Regulator.

3.1 RF Limiter Design

Since sub-micron technology is used for the tag in order to operate at the lowest supply voltage possible, the IC may be damaged if high voltage levels are allowed to appear at the input terminals. Therefore an RF Limiter is used to reduce the AC

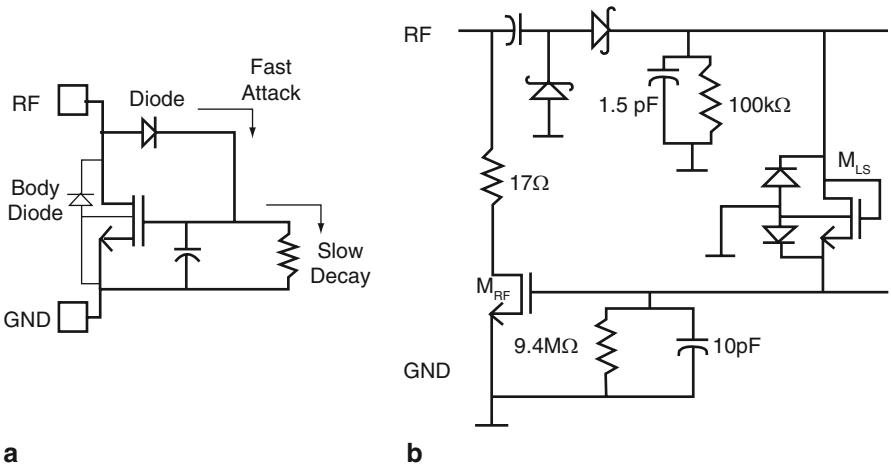


Fig. 8 Example limiter circuits used in RFID. **a** Simple RF limiter. **b** Example RF limiter for 130 nm CMOS

voltage levels at the antenna terminals. For HF designs this is accomplished by detuning the input resonant network, while for UHF designs, this is accomplished by impedance mismatching the antenna. The circuits to perform this function typically use a MOS device as a switch, in series with either a resistor or a capacitor, for mismatching or detuning. One important aspect of the limiter is it must not limit voltage when the input power is low, or it will compromise the sensitivity and reduce read range. Example limiting circuits are shown in Fig. 8.

Figure 8a shows a simple RF Limiter. The Body Diode between the drain and bulk limits negative excursions of the RF signal to about 0.7 V. For positive RF voltage excursions, the NMOS device reduces the input voltage level by providing a low impedance path to GND. The diode connected between the drain and gate of the NMOS transistor provides a fast attack turn on of the NMOS device when the RF pad voltage exceeds the NMOS threshold voltage plus the diode turn on voltage, $VT_{NMOS}+V_D \sim 0.7+0.7=1.4$ V. This occurs as some intermediate input power level. For maximum input power, the V_{GS} of the NMOS device increases and for the given design is about 1.2 V. Thus the maximum positive voltage excursion at the RF pad is about $V_{GS}+V_D \sim 1.2+0.7=1.9$ V. For 130 nm CMOS, the circuit of Fig. 8a is not suitable, because of the large maximum excursion voltage. An improved version of the RF Limiter suitable for 130 nm CMOS is shown in Fig. 8b [11]. A single-stage Schottky diode rectifier is used along with a MOS diode, M_{LS}, and the limiter device M_{RF}. The RF peak voltage at the input pad where limiting transistor M_{RF} turns on is given approximately by

$$2\left(V_{in,peak} - V_d\right) = V_{GS,MRF} + V_{GS,MLS}. \tag{4}$$

Using typical values for V_d and V_{GS}, the value of the peak RF voltage for the limiting action to start is around 1 V, and the value of the peak RF voltage at which the limiting action is very strong is around 1.25 V. Thus the voltage range from the start of limiting to strong limiting is more controlled, taking place over a 0.25 V. This range of operating voltages is suitable for 130 nm CMOS and additional details for the limiter are presented in [11]. RF Limiting is the first line of defense in protecting the IC from over-voltages at high RF power. The next line of defense is the DC Limiter.

3.2 DC Limiter Design

The DC Limiter is used at the output of the rectifier to limit the rectified voltage to an appropriate level by diverting high rectifier output currents to ground. The limiter in some technologies could be as simple as a clamping Zener diode. In state of the art low voltage CMOS processes, the limiter can be quite challenging as it has to work over a very narrow output voltage range, over large current surges out of the rectifier, and over a temperature range typically from –40 to +85°C. The Limiter is complicated by the constraint that when low RF power is available, the limiter must not divert current, or it will impact sensitivity and shorten the read range. Figure 9 shows several DC Limiter architectures.

Figure 9a shows the simplest DC limiter, a Zener diode. This is an acceptable limiter if the subsequent regulator can handle large voltages at the regulator input. Typical clamping voltages for a Zener diode are around 5 V, and for most deep submicron technologies this is too high a voltage for the subsequent regulator to tolerate. Figure 9b shows a limiting clamp made from a string of diodes. This architecture may be acceptable in certain technologies as well, however is does suffer from large clamping voltage variations over temperature and clamping current. The more diodes used the larger the variation. The minimum number of diodes is chosen by taking into account the tags minimum operating voltage, the highest

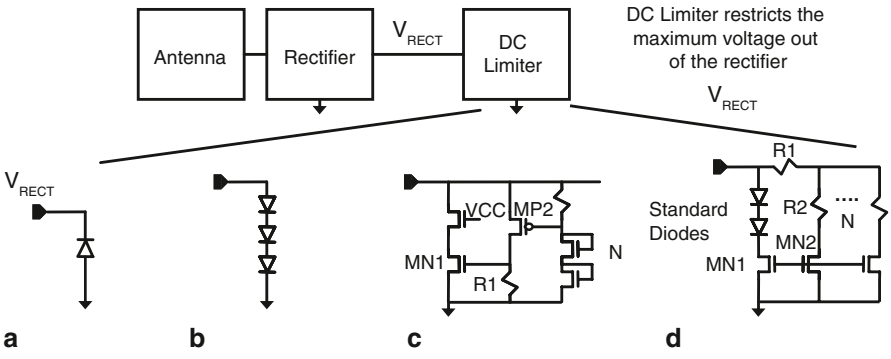


Fig. 9 DC Limiters used in RFID. **a** Zener diode. **b** Multiple forward biased diodes. **c** LF transistor based design in [2]. **d** Low voltage UHF based design

temperature (diode turn on voltage is lowest), the highest process saturation current (strong diode model), and the maximum acceptable off current. These conditions are used to determine how many diodes are required for the clamp to be off during the weak RF power case. Using the minimum number of diodes, if the maximum RF power is presented and the output level is still clamped to a level tolerable by the subsequent regulator, the diode string DC limiter will suffice. In some cases, the diode string will not meet these conditions, and a more complicated DC Limiter will be required. Figure 9c shows the DC limiter used in [2]. The turn on voltage is set by the threshold of MP2 plus the sum of the thresholds in the N "diode connected" NMOS string. Once the input exceeds the sum of the thresholds, the drain current of MP2 is multiplied by resistance R1, resulting in a gate voltage that turns on MN1. MN1 turns on strong pulling current away from the input node, thus creating the strong DC limiting action once the turn on voltage is reached. Other DC limiters are published using a similar principle [6]. Finally, Fig. 9d shows the simplified circuit of the architecture used in [12, 13]. This architecture uses a diode string into a MOS mirror and saturating resistor array. The string of mirrors MN2, R2 etc. act to progressively turn on saturating current sources. The drain resistor R2 is used to saturate the current out of MN2, so the current does not keep increasing with increasing mirror input current. Using this method, the output voltage can be designed to remain relatively flat with increasing input current. This method is useful for low voltage processes where tight control is required to the output voltage without resorting to a complex regulation scheme. The purpose of the DC limiter is to crudely regulate the DC output voltage of the rectifier. Following the DC limiter is the fine voltage regulator which is discussed next.

3.3 Regulator Design

It is desirable in most RFID applications to use the lowest possible supply voltage to operate the tag circuits. Using the lowest supply voltage will result in the highest sensitivity or longest read range for a given reader power and IC current consumption. In current state of the art tags this translates into about 0.8–1 V at the regulator output. Series regulators require some type of device which will incur DC voltage drop from the regulator input to output. Shunt regulators may be implemented as well, and in this case, no voltage loss occurs from input to output. However, shunt regulators present design challenges in stability and may require excess current from the rectifier and DC Limiter output. In either case loss occurs, whether voltage or current, and tradeoffs should be studied for the chosen semiconductor technology. Figure 10 summarizes detail on the series regulator presented in [13]. The series pass device Mpass operates as a switch for input voltages less than 1.45 V. The error amplifier is comprised of a pseudo-bandgap reference using offset MOS devices operating in sub-threshold. A PTAT voltage is forced across R3 via the feedback loop using the 4:1 ratio MOS devices, MN1 and MN2, operating in the sub-threshold region. The PTAT voltage is multiplied up using resistors R2 and R1

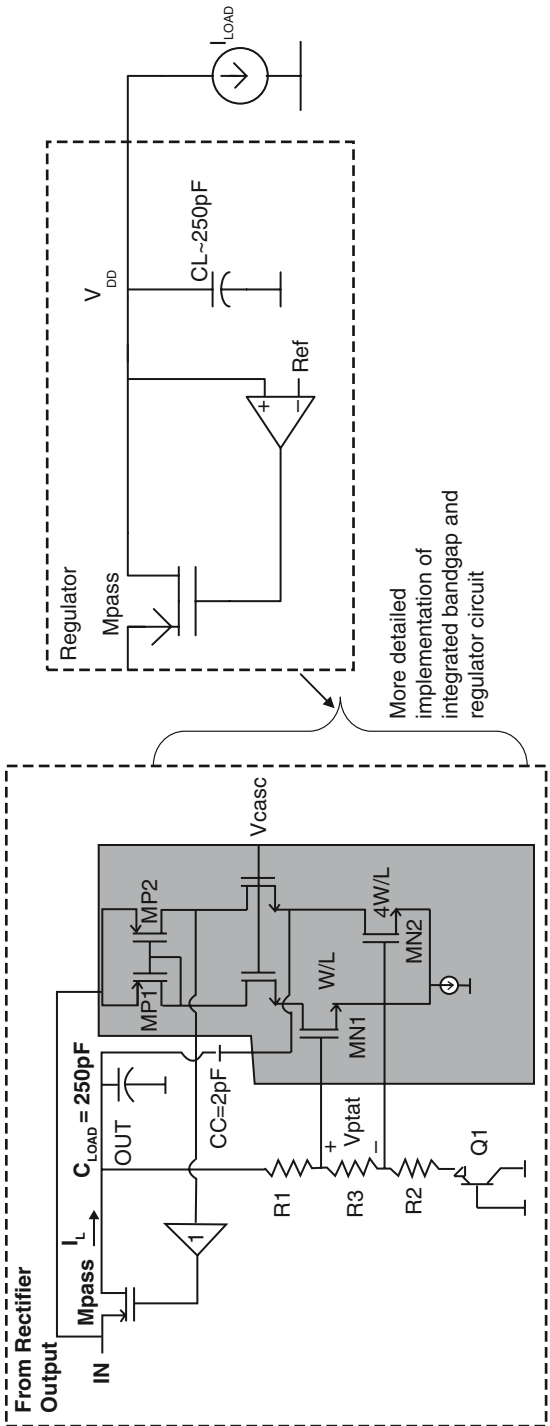


Fig. 10 Regulator architecture used in [13]

and added to the substrate PNP "diode" voltage in the appropriate weighting so as to achieve near zero TC output voltage, i.e. close to the bandgap voltage.

The output voltage in this design is actually set higher than the bandgap voltage, typically 1.2 V, to allow for regulating to higher voltages. In this case the output is regulated near 1.45 V. This results in a slight positive TC for the output regulation voltage and is tolerated by the IC processing circuits. Cascode compensation using Cc = 2 pF is used to guarantee loop stability over a wide range of input voltages and load currents, while using an on chip 250 pF load capacitor. Further details of this regulator circuit are given in [13].

4 RFID Data Receiver Design and Additional Support Circuits

The RFID reader to tag modulation format that is most widely used is amplitude modulation. Amplitude Shift Keying (ASK) is used since the receiver circuits are relatively simple to implement compared to other modulation formats. There are variants of standard ASK, such as Phase Reversal ASK (PR-ASK) or On-Off Keying (OOK) that are used in various applications [1, 14]. The minimum voltage levels in a UHF RFID system are on the order of a few hundred mV at the antenna terminals. Taking into account variants of ASK and modulation depths, the received signal envelope may have only 20 mV difference between logic "1" and logic "0" amplitude levels. The receiver must be able to recover the digital data from this low amplitude envelope signal, and in RFID, a major challenge is to consume as little power as possible, with typically less than a micro-watt of power available for the receiver function. The receiver circuits are comprised of demodulator or envelope extraction circuit, and a data slicer, to convert the demodulated waveform to a single bit digital data stream to be processed by the subsequent digital state machine. HF designs typically extract the system clock from the incoming carrier at 13.56 MHz. UHF systems include a dedicated system oscillator to create the proper timing signals for the receiver, transmitter and digital processor. This section will outline the UHF RFID data receiver design in more detail.

4.1 Data Receiver Design

Simple demodulation of ASK signals is performed using envelope extraction, threshold estimation and envelope to threshold comparison. The resulting single bit digital data stream is sent to a digital data detector for recovering the encoded data bits. Figure 11 shows the RFID receiver circuit used in [11].

The figure shows the antenna equivalent circuit, diode doubler used for envelope extraction, Variable Gain Attenuator, weighted average circuit (Rb, Rav and Cav), and data slicer comprised of a preamp and dynamic latch. The input to the variable

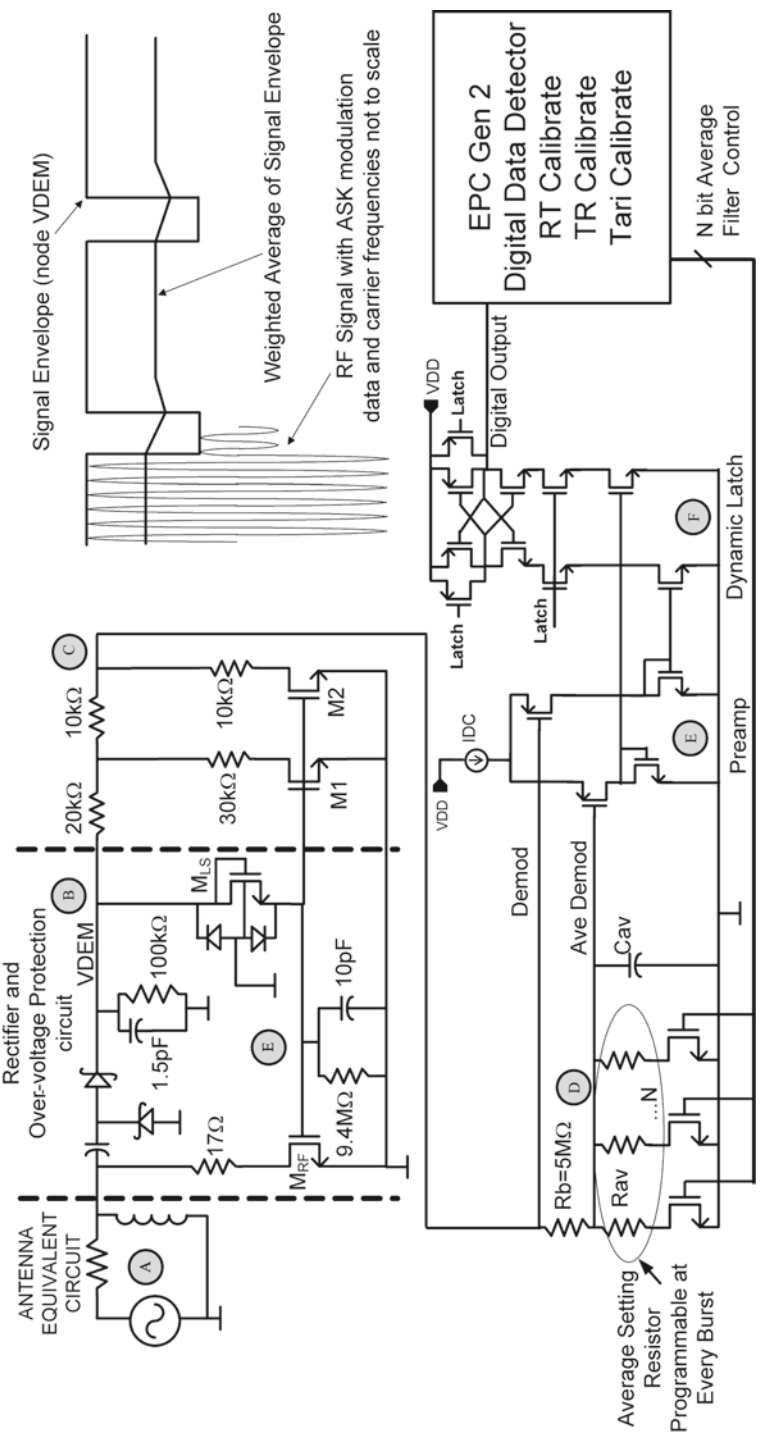


Fig. 11 RFID ASK receiver circuit used in [11]

attenuator is the demodulated signal at node B. The wide dynamic range of the RF is limited by the previously described RF limiter loop. The attenuator further deals with the dynamic range by reducing the common mode input voltage to the data slicer. This is required as the chosen slicer architecture uses a DC coupled input and the comparator cannot tolerate a high common mode input voltage. A weighted average circuit is used in place of a typical averaging filter. For a typical averaging filter Rav is removed from Fig. 11, and a low pass filter consisting of Rb and Cav is used to estimate the threshold. This method works fine for DC free codes such as Manchester encoded data, where the high and low levels of the demodulated waveform are equally likely. Since the modulation used in the EPC Gen two protocol is not DC free, a weighted average is used estimate the threshold voltage. The data-slicer compares the demodulated envelope, node C, to its weighted average, node D, formed by Rb, Rav and Cav. The purpose of Rav, is to weight the average towards zero in the presence of long periods of high levels of demodulated data such as those presented by EPC Gen 2 PR-ASK. Rav is a digitally selectable resistor chosen during the data preamble, and more details on this implementation given in [11].

4.2 Data Slicer Design

The slicer converts the demodulated analog signal to a single bit digital data stream using an over sampled clock of 1.28 MHz. Therefore the final digital output data follows the envelope of the RF with small timing shifts due to the free running over sampled clock. The circuit used in [11] replaces a standard high gain analog comparator and DFF, with a preamp and dynamic latch similar to what is used in many A to D converters. The small input differential voltage is pre-amplified and then further amplified to full CMOS levels with a latching signal and strong positive feedback. Figure 12 shows the preamp and dynamic latch in block and schematic form.

The preamp, comprised of M0, M1, M2, M3 and M4, is required to reduce the large input referred offset voltage of the dynamic latch, M5 through M14. The advantage of the dynamic latch is it consumes no static power, and it has a very large gain, implemented using strong positive feedback. The preamp is constantly powered by the current source M0, while the latch is dynamically powered at the resolution event. The function of the latch is to sense the small difference current created by the preamp and current mirrors, M3 through M6, and resolve the current difference to a full CMOS level. Referring to Fig. 12c, when the latch logic signal is low, the two output voltages are reset high. As the latch logic signal transitions high, the resolution period begins. This is labeled the “HIGH CURRENT RESOLVE TIME”. During this time period both outputs transition to a common mode voltage and the small delta current out of M5 and M6 produces a small delta voltage at the output. The strong positive feedback drives the output to transition in the direction favored by the input voltage, thus producing the final digital output value. The single bit digital data stream is then decoded by the subsequent all digital clock and data recovery circuit.

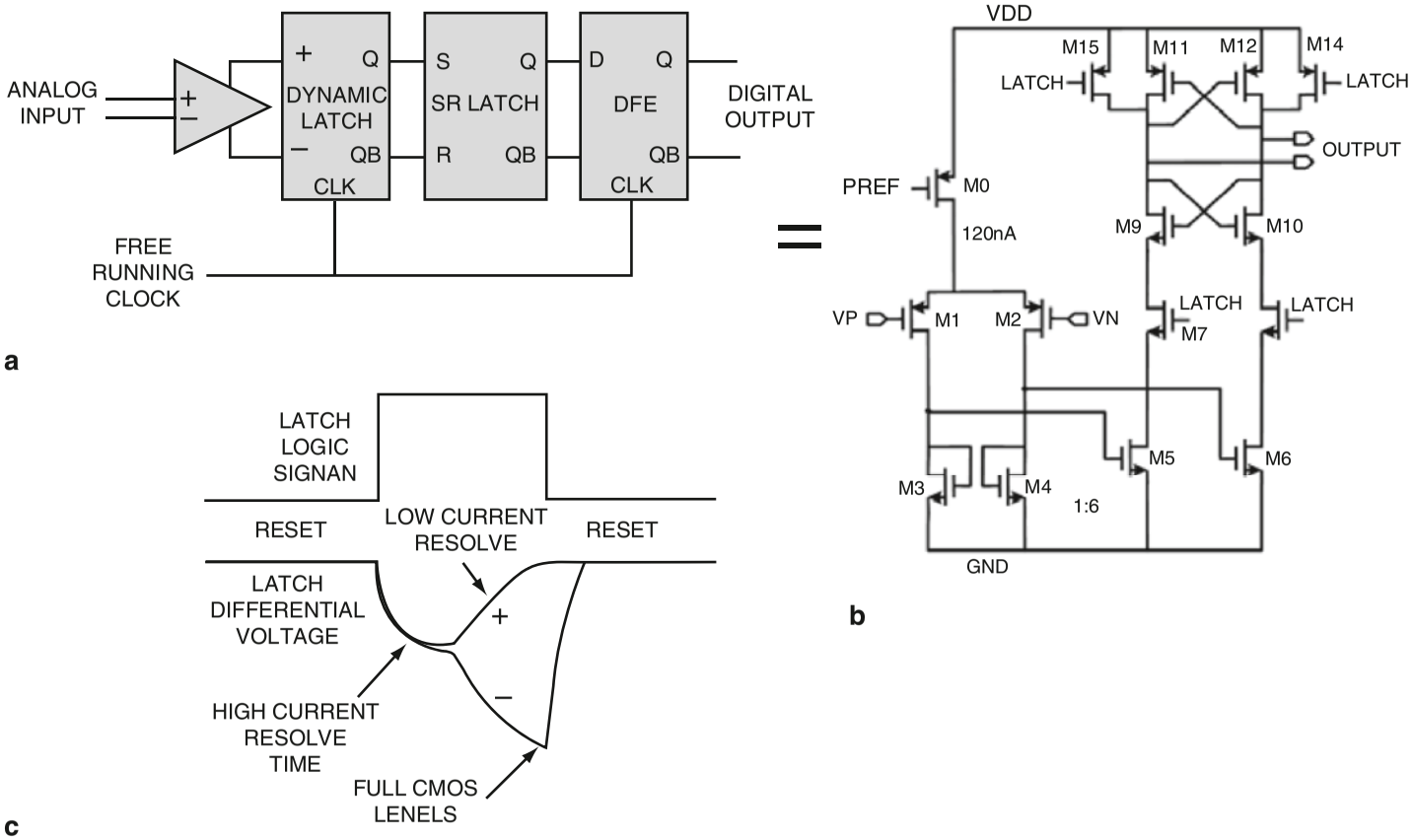


Fig. 12 RFID demodulator slicer. **a** Dynamic latch block diagram. **b** Dynamic latch front end schematic. **c** Differntial logic output of dynamic latch

4.3 Additional RFID Circuits

Additional important RFID circuits include the system oscillator [6, 15], the Power On Reset (POR) circuit [6], the EEPROM controller [16], and random number generator [17]. As the scope of this paper is limited, the reader is referred to the listed publications for details on implementations of these circuits.

5 Flexible Organic Semiconductor RFID Tags

It is highly desirable that RFID tags used for many applications be flexible. For the most part, the RFID tag using a silicon IC is flexible; however the chip substrate itself is not flexible and can be easily damaged due to mechanical stress. Research is ongoing into RFID circuits that are processed using flexible organic semiconductor technology. In addition to being flexible, this technology has the potential to be very cost effective, a dominating constraint in RFID tags used for bar code replacement. This section overviews a few design issues concerning organic semiconductor RFID tags from a circuit design perspective.

The most widely used organic semiconductor material is pentacene and the corresponding available device is the P-type Organic Thin Film Transistor (TFT) [18, 19]. The mobility of pentacene has improved with research but is still in the sub 1 cm2/V-s range, several orders of magnitude lower than the mobility of silicon. This translates into low frequency of operation; a potential issue when using TFT's in certain applications. It also limits circuit topologies to PMOS only, and this is a challenge in creating low power logic circuits. This is in contrast to using CMOS logic which has the benefit of near zero static current. Figure 13 shows a few circuits used in published organic RFID tags. Figure 13a shows one of the rectifier topologies used in [18], one of the first complete organic tags published in the literature. The structure is relatively simple by necessity as the tag is made in low cost PMOS only technology. The rectifier uses a diode connect PMOS similar to the single diode architecture of Fig. 3a. The double half wave rectifier shown in Fig. 13b was used in [19] as it provides nearly double the output voltage. In this case the process included vertical diodes which perform better than the PMOS diode connected counterpart. Other works also include vertical diode based rectifiers such as the simple single diode rectifier given in [20].

PMOS only design requires the use of specialized logic [18] as shown in the NAND gate in Fig. 13c. The load in this case is the zero Vgs connected PMOS which acts like a high impedance load. Since the PMOS device never turns off, significant leakage exists at zero Vgs. Thus the device is used as a load similar to the all NMOS depletion mode/enhancement mode logic used early on in silicon logic design before CMOS was readily available. In fact many techniques used in the early days of NMOS only design can be reused in organic PMOS design by

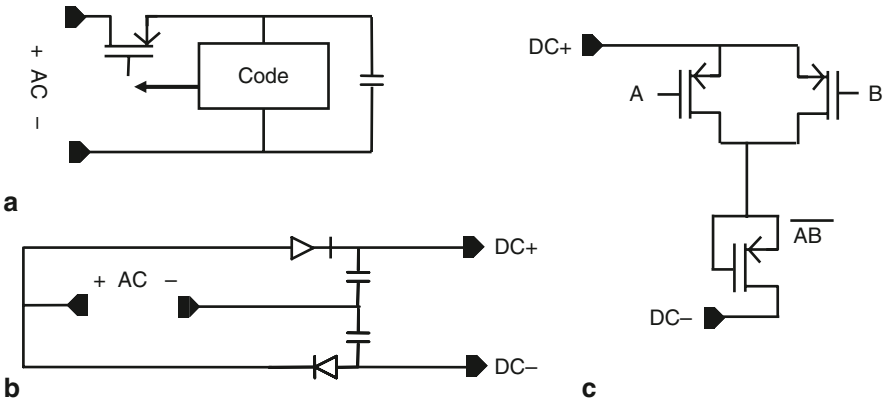


Fig. 13 Organic based RFID Circuits. **a** PMOS only rectifier presented in [18]. **b** Double half wave rectifier presented in [19]. **c** PMOS only NAND gate with zero Vgs load in [18]

inverting the structures. Many additional analog circuits have been demonstrated in PMOS only organic technology. In the past few years, entire tags have been demonstrated in PMOS only technology with promising results. Ongoing research continues in circuit design using PMOS only devices. In addition, research is ongoing to find a cost effective way to add NMOS to the flexible organic circuit, while keeping cost an absolute minimum. Having organic CMOS to design with will make the design of low cost organic tags more similar to their silicon counterparts. Many challenges remain on the processing side of organic tags. A dominant issue with TFT circuits manufactured today is electrical stability; that is the characteristics of the device degrade over time with use. Given these challenges, much research is still required to make flexible RFID tags made from organic technologies a viable market option. Research includes the discovery of new flexible semiconductor materials, materials that support both P-type and N-type transistors, ways to increase the mobility, and innovative circuit techniques to use the cheapest technology available.

6 Conclusions

The passive RFID tag consists of an antenna and IC operating without a battery. The basic functional analog circuits of the IC include rectifier, RF and DC power management, data receiver, backscatter modulator, non-volatile memory controller, and additional sub-circuits. This paper addressed several aspects of the design of RF and Analog Front End circuits for RFID.

References

1. K. Finkenzelle, *RFID Handbook: Radio Frequency Identification Fundamentals and Applications*. (Wiley, New York, 1999)
2. U. Kaiser, W. Steinhagen, IEEE J. Solid-State Circuits. **30**(3), 306–310 (1995)
3. U. Karthaus, M. Fischer, IEEE J. Solid-State Circuits. **38**(10), 1602–1608 (2003)
4. M. Usami, in *IEEE RFIC Symposium Digest* of Papers, Jun 2004, pp. 241–244
5. R. Glidden, C. Bockorick, S. Cooper, C. Diorio, D. Dressler, V. Gutnik, C. Hagen, D. Hara, T. Hass, T. Humes, J. Hyde, R. Oliver, O. Onen, A. Pesavento, K. Sundstrom and M. Thomas, IEEE Communications Magazine. **42**, 140–151 (2004)
6. J-P Curty, N. Joehl, C. Dehollain, M. Declercq, IEEE J. Solid-State Circuits. **40**(11), 1602–1608 (2005)
7. G. De Vita, G. Iannaccone, IEEE Trans. Microw. Theory Tech. **53**(9), 2978–2990 (2005)
8. R. Barnett, J. Liu, S. Lazar, IEEE J. Solid-State Circuits. **4**(2), 354–370 (2009)
9. T. Umeda, H. Yoshida, S. Sekine, Y. Fujita, T. Suzuki, S. Otaka, IEEE J. Solid-State Circuits. **41**(1), 35–41 (2006)
10. J.F. Dickson, IEEE J. Solid-State Circuits. **SC-11**(3), 374–378 (1976)
11. G. Balachandran, R. Barnett, in *CICC 2009 Proceedings*, Sept 2009, pp. 383–386
12. R. Barnett, G. Balachandran, S. Lazar, B. Kramer, G. Konnail, S. Rajasekhar, V. Drobny, in *Proceedings of ISSCC*, San Francisco, Feb 2007, pp. 582, 583
13. G. Balachandran, R. Barnett, IEEE J. Solid-State Circuits. **41**(9), 2019–2028 (2006)
14. EPC class 1 generation 2 UHF air interface protocol standard version 1.0.9. (2005), http://www.epcglobalinc.org/standards_technology/specifications.html
15. R. Barnett, J. Liu, in *CICC 2006 Proceedings*, 10–13 Sept 2006, pp. 769–772
16. R. Barnett, J. Liu, in *Proceedings of IEEE CICC*, Sept 2007, pp. 393–396
17. G. Balachandran, R. Barnett, IEEE Circuits. Sys. I. **55**, 3723–3732 (2009)
18. E. Cantatore, T.C.T. Geuns, G.H. Gelinck, E. van Veenendaal, A.F.A. Gruijthuijsen, L. Schrijnemakers, S. Drews, D.M. de Leeuw, IEEE J. Solid-State Circuits. **42**(1), 84–92 (2007)
19. K. Myny, S. van Winckel, S. Steudel, P. Vicca, S. De Jonge, M.J. Beenhakkers, C.W. Sele, N.A.J.M. van Aerle, G.H. Gelinck, J. Genoe, P. Heremans, *ISSCC Digest Technical Papers*, Feb 2008, pp. 290, 291
20. Y. Ai, S. Gowrisanker, H. Jia, I. Trachtenberg, E. Vogel, R.M. Wallace, B.E. Gnade, R. Barnett, H. Stiegler, H. Edwards, Appl. Phys. Lett. **90**, 262105 (2007)

A Dual Frequency Band Comprehensive RFID TAG

Albert Missoni, Günter Hofer and Wolfgang Pribyl

1 Introduction

Several RFID systems in a diversity of ISM frequency bands exist. These systems can be distinguished by the maximum operation distance, transponder or reader infrastructure costs, communication speed or security features [1]. The developed transponder should be able to be accessed very precisely in an area where also several other transponders are close to it. At the product initialization phase and at the point of sale transponders are close coupled. The comprehensive transponder should also provide a stable contactless communication when commodities are placed sparse on a palette or when tagged apparels in racks should be detected during an inventory. To fulfill these preconditions far field communication over several meters has to be supported. In general the transponder has to fulfill two target application segments. The first segment is the so called item identification and the other segment is called transport and packaging. Nowadays both segments differ in the established RFID frequency and in the communication protocol.

According to the IDTechEX Research Group the segment transport and packaging in form of tagged pallets and boxes generated market revenues of $ 225 million in the year 2009. The preferred contactless communication for transport and packaging is the ISO18000-6 UHF standard. In the near future this standard will be detached by the UHF EPC Gen 2 specification [2]. In the second application segment we have to support—the item identification—the HF standard ISO15693 is well established.

Finally we decided to support the UHF EPC Gen2 and a pre-release of the EPC HF standard [3]. Especially because of synergies in the digital processing unit these two protocol standards have been chosen for contactless communication. With a simple firmware update a modern ISO15693 reader can be prepared to communicate also with EPC HF transponders.

A. Missoni (✉)
Infineon Technologies Austria AG, Graz, Austria

A. Missoni
Graz University of Technology Institute of Electronics, Graz, Austria

H. Casier et al. (eds.), *Analog Circuit Design,*
DOI 10.1007/978-94-007-0391-9_16, © Springer Science+Business Media B.V. 2011

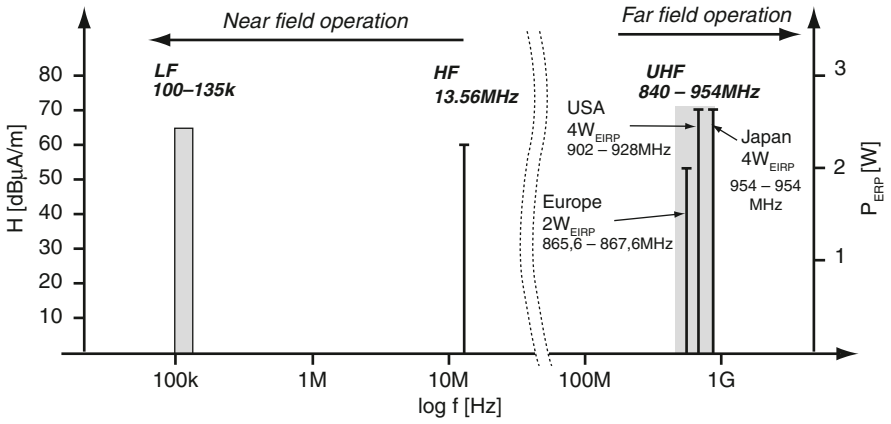


Fig. 1 The most relevant frequency bands for RFID applications

In general with the Electronic Product Code (EPC) it is possible to identify every object with a ubiquitous and globally unique code. Attached to a worldwide network and compatible software infrastructure the so called Global Data Synchronization (GDS) every EPC object's appearance can be written into the database or the object information can be read from the database [4].

In Fig. 1 the ISM RFID bands most relevant for market revenue are shown. In the LF and HF band coupling between reader and transponder relies on the magnetic field (inductive coupling). The magnetic field intensity is described with the magnetic field strength H. Coupling in the UHF band relies on the radiation field (electromagnetic coupling) and the emitted power P_{ERP} describes the intensity. In the UHF frequency band multiple center frequencies, operation bandwidths and emitted power levels are specified worldwide. Special antenna design measures are necessary to carry the transponder quality factor Q on a high level over this very broad RFID UHF bandwidth [5].

Transponder operation distance is in HF near field operation well defined to approximately half a meter with reader antennas of a diameter of 30 cm. Typical UHF low power tags are designed for far field operation and can reach a distance of about 5 m. A big distance could be a disadvantage when privacy is desired during the contactless operation. Expensive shielding methods or multiple reader systems with distance estimation are necessary when a reliable short range operation has to be guaranteed in a UHF system. Material sensitivity of radiated waves could also be a reason to go either for UHF or HF–RFID systems.

A transponder can be divided into two parts—into the antenna and into the chip. The simplified circuit of these elements is shown in Fig. 2. Passive coil or antenna components and the non linear chip impedance are concentrated in one inductance and in one capacitance.

The following equations [1, 6, 7] will give an overview how to calculate the chip input voltage u_{chip} in the HF and UHF field from the national specified magnetic field strength H (A/m) or from the specified emitted power P_{ERP} (dBm) which are shown in Fig. 1.

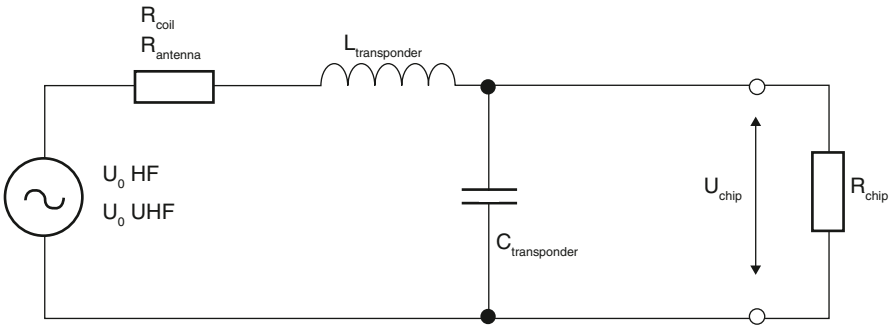


Fig. 2 UHF and HF equivalent resonance circuit

Inductive and capacitive coupling are usually near field powering methods. Biot-Savart's law can be taken to describe the magnetic field emitted by the reader:

$$\vec{H}(\vec{x}) = \frac{i_1}{4\pi} \cdot \oint_S \frac{\vec{dl} \times \vec{x}}{|x|^3} \tag{1}$$

In Eq. 1 the magnetic field strength is defined by the current i_1 through the current loop S and by the distance x between the loop and the observation point. Cite(Leuchtmann) It can be shown that the Biot-Savart's law results in a magnetic field reduction according x^3 or 60 dB per distance of decade. Using the magnetic flux density B of Eq. 2 the induced voltage in the transponder coil can

$$B = \mu_0 \cdot \mu_r \cdot H \tag{2}$$

be defined as shown in Eq. 3.

$$u_{0\,HF} = N_2 \cdot \frac{d}{dt} \int_A B\, dA \tag{3}$$

N2 describes the number of transponder coil turns and A the medial coil area. As a resonance circuit is used, the induced voltage $u_{0\,HF}$ will be boosted by the quality factor Q. A general quality factor expression is shown in Eq. 4

$$Q(\omega) = \frac{1}{2\pi} \cdot \frac{energy\ stored}{energy\ dissipated} = \frac{1}{2\pi} \cdot \frac{E_{magnetic} - E_{electric}}{E_{loss\ per\ cycle}} \tag{4}$$

With the quality factor Q the relation between the induced voltage and the chip input voltage can be derived.

$$u_{chip} = u_{0HF} \cdot Q_{transponder} = u_{0HF} \cdot \frac{\omega_r \cdot L_{transponder}}{R_{transponder}} \tag{5}$$

The transponder resistance $R_{transponder}$ comprises the coil resistor R_{coil} or the antenna resistor $R_{antenna}$ and the non linear chip resistor R_{chip}.

In UHF systems maximum power levels P_{ERP} or P_{EIRP} are defined in each country. The following simplified equations show how to get from a specified power level to the according chip input voltage. The radiation density S of an electromagnetic field is defined by the Poynting vector in Eq. 7 as a vector product of electric and magnetic field.

$$\vec{S} = \vec{E} \times \vec{H} \tag{6}$$

With the help of the isotropic radiated density of Eq. 7 a correlation between

$$S = \frac{P_{EIRP}}{4\pi \cdot x^2} \tag{7}$$

power level P_{EIRP} and the electric field strength E can be expressed Eq. 8.

$$E = \sqrt{S \cdot Z_F} \qquad Characteristic\ wave\ impedance\ Z_F = 377\Omega \tag{8}$$

Compared with the HF system where the magnetic field strength drops with 60 dB, the biggest benefit for a far field operating UHF system can be extracted from Eq. 9. The electric field drops in far field proportional to the distance x which corresponds to a loss of 20 dB per distance decade.

$$E \approx \frac{\sqrt{30 \cdot P_{EIRP}}}{x} \tag{9}$$

One parameter of a matched radiation antenna is the effective length l_0. With Eq. 11 an estimation for the open circuit voltage $u_{0\ UHF}$ is possible.

$$u_{0\ UHF} = \int_l E\ dl \approx l_0 \cdot E \tag{10}$$

This typical small voltage $u_{0\ UHF}$ has to be multiplied with the quality factor of the UHF transponder to get the chip input voltage u_{chip}.

$$u_{chip} = u_{0UHF} \cdot Q_{transponder} = u_{0UHF} \cdot \frac{\omega_r \cdot L_{transponder}}{R_{transponder}} \tag{11}$$

1.1 Transponder Architecture

Either a UHF reader or a HF reader can power this chip via the selective transponder antenna. Switching on both readers simultaneously, the internal frequency band decision unit will optimize the 13.56 MHz power generation unit. HF is preferred

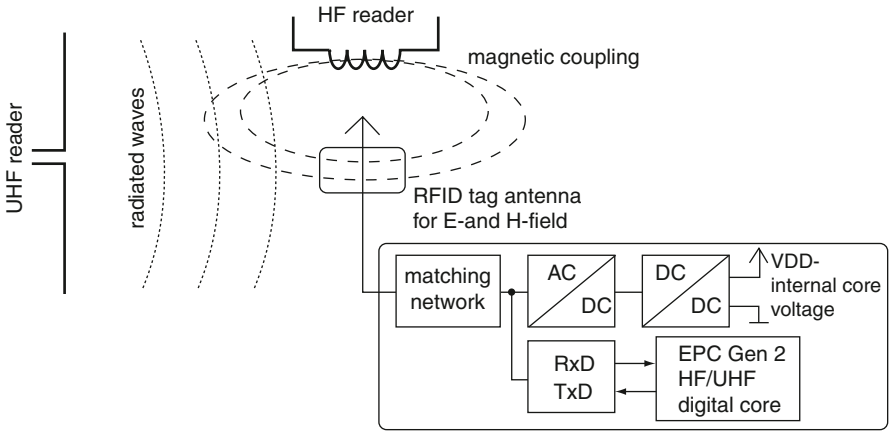


Fig. 3 Comprehensive transponder architecture

because of its reliable reduced maximum communication distance. The customer has to bring the transponder close to the reader and the possibility is high, that the customer would like to communicate with exactly this reader. The architecture of the transponder and analog frontend is shown in Fig. 3. No integrated passive components or even additional external components are necessary for a good antenna to chip input impedance matching. Tasks like blocking the big HF coil inductance and parasitic capacitance from the UHF antenna are done by the dual frequency selective antenna design [8].

The matching network on the chip includes the parasitic capacitance and the equivalent resistive chip load. The alternating sinusoidal waveform is rectified with an AC/DC converter to a signal with 27.12 MHz ripple and either a too small or a too big average voltage level. A DC/DC converter will convert this signal to the right power supply voltage level VDD.

In parallel to the power supply unit the communication modules are attached to the antenna. Two separated RxD modules for receiving information from the reader are used. The TxD unit which includes the powerful shunt transistor can be shared for all frequencies and only small modifications are necessary when the frequency band is changed. A digital EPC encoder, decoder and control unit for the analog interface is also part of the chip.

2 HF and UHF Power Generation with One Antenna Port

A typical UHF rectifier consists of a low voltage single-ended (unbalanced) Dickson or Greinacher charge pump [9–11]. In Fig. 4 one antenna node is connected to chip VSS which is equivalent to substrate (VSS is the negative power supply) and the other to a capacitive multi stage charge pump. For good rectifier power ef-

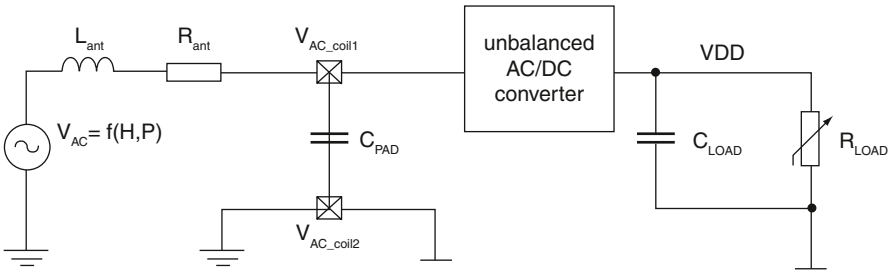


Fig. 4 Single ended rectifier concept

ficiency Schottky diodes with MIM (metal–isolator–metal) or metal–metal capacitors are typically used.

Using MOS transistors for this rectifier concept, capacitors to decouple the chip input stage from the sinusoidal antenna voltage are necessary. The antenna amplitude is symmetrical related to VSS. The negative amplitude would trigger reverse diodes or even parasitic bipolar transistors with a moderate current amplification factor.

For UHF only products the decoupling capacitor C_{AC} value in Fig. 5 is typically below 1 pF. At about 900 MHz the impedance of the coupling capacitor is low and the peak to peak voltage at the rectifier is similar to the amplitude at the antenna node. The parasitic PN diode or the MOS transistor Mn2 is defining the negative voltage level with its forward voltages. Transistor Mn1 is the rectifier diode of the voltage doubler circuit.

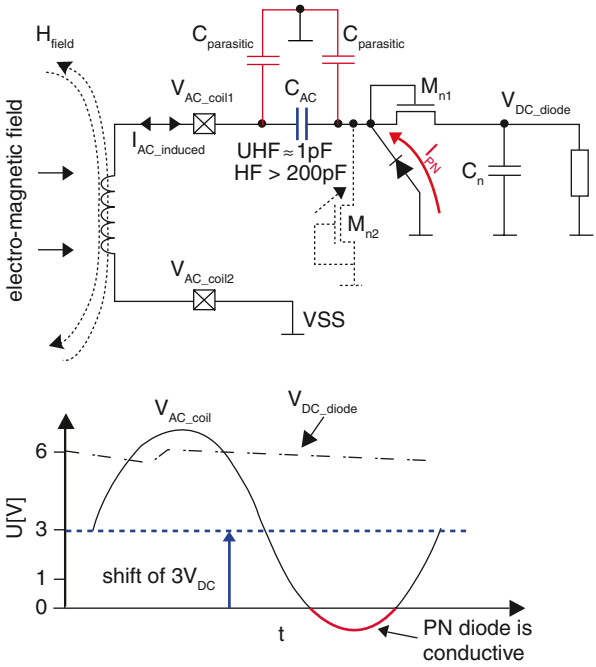


Fig. 5 Single ended rectifier with reverse diodes

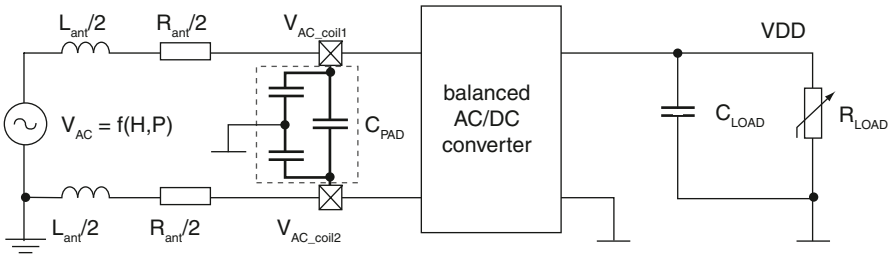


Fig. 6 Balanced rectifier structure

Should this concept also work in the 13.56 MHz HF near field, a capacitor value of several hundred pF is necessary. Besides the big charge pump capacitor C_{AC} also additional parasitic capacitors $C_{parasitic}$ to the lossy chip substrate will occur. Voltage amplification factor Q and the rectifier efficiency will be reduced. The equivalent chip input capacity will be too big for a typical inductance used in a HF RFID resonance circuit. Eq. (12) shows the impact on the chip input voltage of a conjugate matched input chip circuit which is operated in the UHF band [12].

$$\frac{u_{chip}}{u_{0\,UHF}} \approx \frac{1}{2\omega \cdot C_{chip\ serial} \cdot R_{chip\ serial}} \tag{12}$$

If an equivalent serial chip resistance $R_{chip\ serial}$ of 30 Ω is assumed and the serial chip capacitance $C_{chip\ serial}$ is UHF product typical 500 fF, an amplitude amplification factor of about 6 can be expected. Is the capacitor because of HF operation drastically increased to C_{AC}=300 pF and a parasitic coupling capacitor value to VSS of about $C_{parasitic}$=30 pF expected, we get instead of a voltage amplification a voltage damping factor of 10. The conclusion is not to use an AC coupling capacitor rectifier structure when the transponder has also to work in the HF RFID band.

Input structures without AC coupling are used for typical HF passive RFID input circuits. In Fig. 6 a balanced rectifier with symmetrical antenna terminals is shown. A cross coupled NMOS transistor pair will shift the coil voltage above VSS (see Fig. 10). No negative voltages will occur.

No parasitic reverse PN—diode will conduct and no AC coupling capacitor is necessary. A drawback of this concept is the inefficient rectifier at the analog frontend which delivers a DC voltage which is typically lower than the input peak voltage [12, 22]. To replace the unbalanced UHF RF charge pump at the antenna to chip interface [13, 14], a low drop rectifier and a low voltage DC/DC up converter are necessary and they are introduced in the following chapters.

2.1 Power Generator Architecture

In Fig. 7 the power generation architecture is shown [15]. At the balanced rectifier structure the cross coupled NMOS transistor pair Mn1 and Mn2 generates the

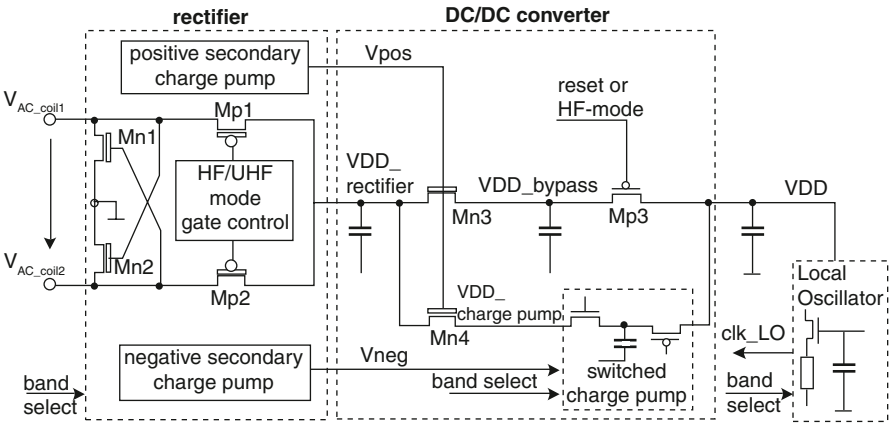


Fig. 7 Power unit architecture

internal VSS potential. Once in a period both antenna nodes V_{AC_coil1} and V_{AC_coil2} are shorted for a very short moment to the internal VSS potential. This guarantees a minimum voltage potential at the two antenna nodes of about −150 mV. The NMOS transistors should conduct well because in the low power UHF mode each negative mV below VSS will reduce the positive peak voltage and therefore also the power unit efficiency. Also attached to the alternating antenna nodes are the rectifier PMOS transistors Mp1 and Mp2. These transistors are controlled in a different way when HF or UHF field is applied. Two secondary low power charge pumps are also located in the rectifier. They are designed to deliver a reliable output current of up to 100 nA. With the help of these secondary charge pumps the two attached DC/DC converters can be operated even with input voltages of 0.6 V and high efficiency. During the startup phase when HF or UHF field is applied to the chip, a power bypass mode is active to charge up the internal power supply VDD.

2.2 Multi Frequency Rectifier

A dual frequency rectifier is shown in Fig. 8. In the UHF mode, the PMOS transistors Mp1 and Mp2 in the rectifier are controlled by capacitive coupling of the two capacitors C1 and C2. As the voltages at both antenna nodes are 180° phase shifted, the biggest amplitude at Mp1 and Mp2 will occur during the positive and negative peak voltage phases of $V_{AC\text{-}coil1}$ and $V_{AC\text{-}coil2}$. Switches S1 and S2 are permanently closed in the UHF mode. Therefore the two NMOS transistors Mn5 and Mn6 will define the high impedance capacitor nodes to VSS when the according coil voltage terminals are higher than one threshold plus overdrive voltage of about 200 mV. When the positive coil peak voltage at node $V_{AC\text{-}coil2}$ is reached, the gate potential of Mp1 is defined well by Mn6 to VSS. In the next phase the $V_{AC\text{-}coil2}$ potential will drop close to VSS and the potential at node $V_{AC\text{-}coil1}$ will increase to its positive

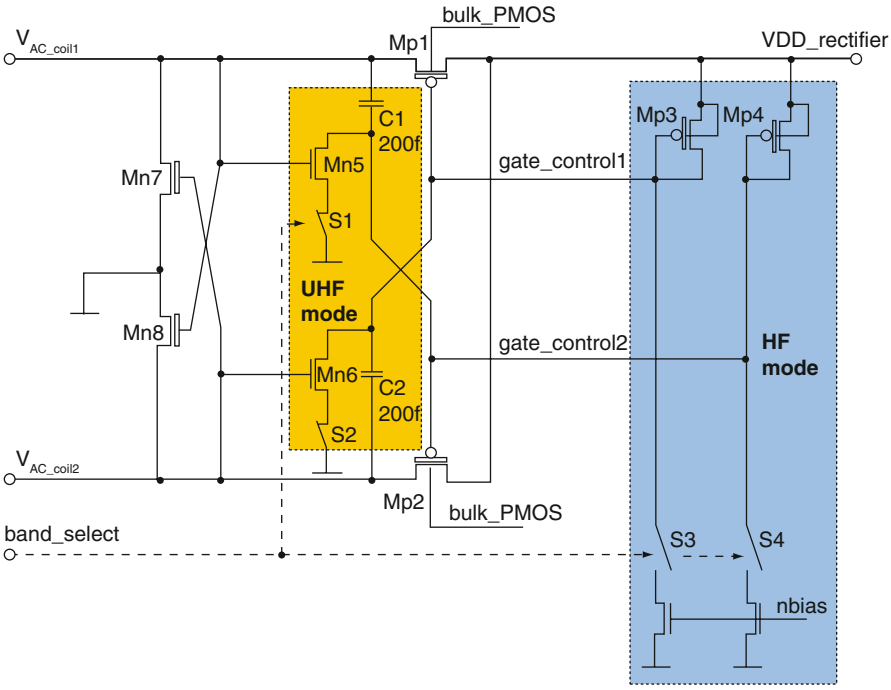


Fig. 8 Dual frequency rectifier

peak voltage. Is the coil voltage $V_{AC\text{-}coil2}$ only little above the threshold voltage of Mn6, the coupling will be stronger than the resistive part of Mn6. The high impedance node of capacitor C2 is coupled to a negative potential which is similar to the threshold voltage of Mn6. On the other side the more positive coil voltage potential will bring the rectifier transistor Mp1 into a conductive state. Current will be transported from the antenna node to the internal rectified node *VDD_rectifier*.

In the HF mode the coil voltage is stabilized to about 4 V_{peak} to support sideband voltage requirements defined in the ISO10373. Because of this requirement the low voltage UHF control-mode cannot be used. The coil voltage would be clamped by the rectifier at a much too low voltage level of about 1.8 V_{peak} in moderate filed strength conditions. To change into the HF mode, the switches S1/S2 have to be opened and S3/S4 closed. The gate terminals of the rectifier PMOS transistors Mp1 and Mp2 are connected now to two PMOS diodes Mp3 and Mp4 [16]. As node *VDD_rectifier* is a stable DC voltage, the control voltages *gate_control1* and *gate_control2* are also stable DC potentials. Mp1 and Mp2 will only conduct well around the positive coil peak voltage and no current clamping mechanism will occur.

During the startup phase signal *band_select* is at its power on reset value. In this case the rectifier is in the UHF mode. The coil voltage is clamped in this mode to about 1.5 V_{peak} which is enough to detect the frequency and to define the *band_select* signal.

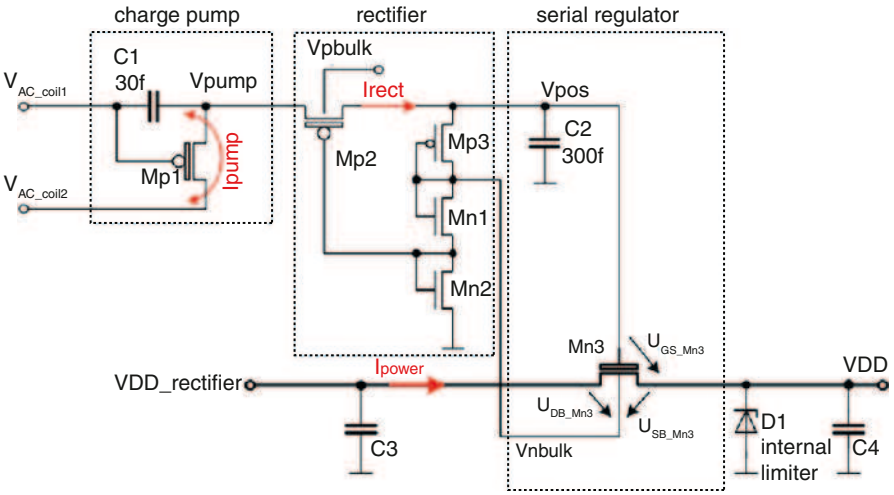


Fig. 9 High voltage blocker as DC/DC down-converter

2.3 DC/DC Down Converter

Two DC/DC down-converters are necessary to reduce during the HF mode the rectified *VDD_rectifier* voltage of 3.5 V down to a VDD power supply voltage level of 1.4 V which is necessary for a modern 120 nm logic. On the other hand in the UHF mode a down converter is not necessary and the voltage drop should be as small as possible. In the architecture overview of Fig. 7 these regulators are represented by NMOS transistors Mn3 and Mn4.

In Fig. 9 the DC/DC converter in form of a NMOS source follower circuit is shown. In the HF mode the positive charge pump defined by capacitor C1 and PMOS transistor Mp1 would double the already high coil potential to about 7 V. C1 with a value of 30 fF is rather weak at the carrier frequency of 13.56 MHz and instead of a powerful voltage doubler we just get a weak serial resistor between the coil node and the rectifier diode Mp2. Only when the potential of *Vpump* is above the threshold voltage of Mn2 and Mp3, charge will be transferred via Mp2 to the serial regulator gate node of Mn3. *Vpos* is stabilized with the 300 fF capacitor C2 and defined by three threshold voltages of transistors Mp3, Mn1 and Mn2. A typical value for *Vpos* is 2.5 V which can be taken from Fig. 10. A big advantage of using a NMOS transistor is the very good PSRR which can be achieved with very low power efforts.

The current *Irect* varies with the coil voltage and reaches an average value of 150 nA at 4 V_{peak}. In the UHF mode not only a weak positive but also a weak negative charge pump *Vneg* is delivered by the rectifier.

In Fig. 11 the small voltage drop from *VDD_rectifier* to *VDD_chpump* along the serial regulator is shown.

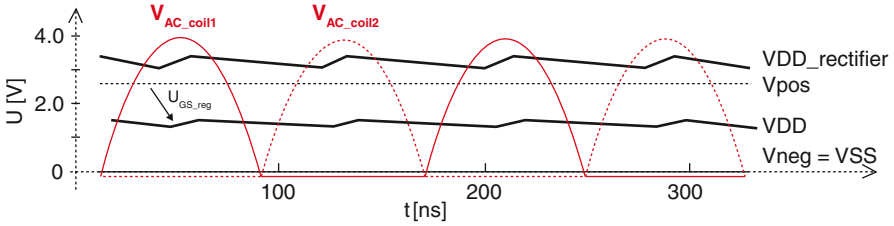


Fig. 10 Transient timing of the power path in the HF mode

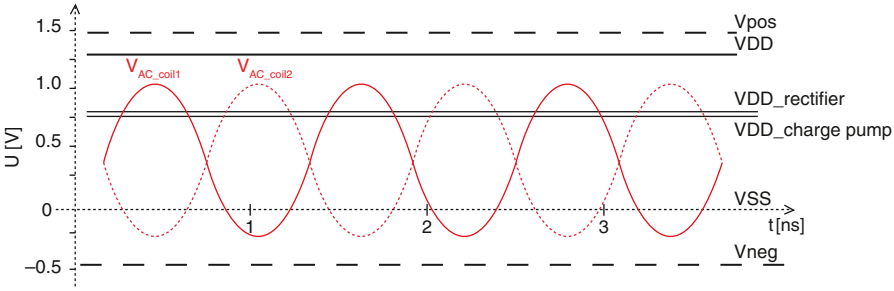


Fig. 11 Transient timing of the power path in the UHF mode

2.4 DC/DC Up Converter

In the UHF far field the power at the rectifier drops with 20 dB per distance decode much smoother than in the HF near field. An additional effort in power generation optimization will result in a significant increased far field operational distance. Typically the charge pump is directly powered by the coil antenna nodes an operated with the UHF frequency. In this work the low drop rectifier and the serial regulator delivers the potential *VDD_chpump*. Additionally, the weak negative charge pump delivers about –400 mV to increase the control voltage potentials at *MN1_gate* and *MP1_gate* in Fig. 12.

This low voltage charge pump can operate already at 550 mV input voltage with a power efficiency of 69% at 8 µA average load at the internal power supply node VDD. Capacitors C2–C4 with a value of 50 fF are part of the control signal generation for the main power charge pump represented by transistor Mn1 and Mp1. We expect an incoming power of –10 dBm at the rectifier which enough to generate a typical *VDD_chpump* voltage of 600 mV. During the ON phase of Mn1 a potential of 1.3 V is applied to the gate. This will guarantee that the voltage drop over Mn1 is minimal and the potential at C1 is at the end of the charging phase similar to *VDD_chpump*. Only with the help of the negative secondary charge pump this voltage level is possible—else we would have to expect about 0.95 V at the gate of Mn1. In the second phase (which is defined by the incoming clock *f_osc*) the gate of Mn1 is

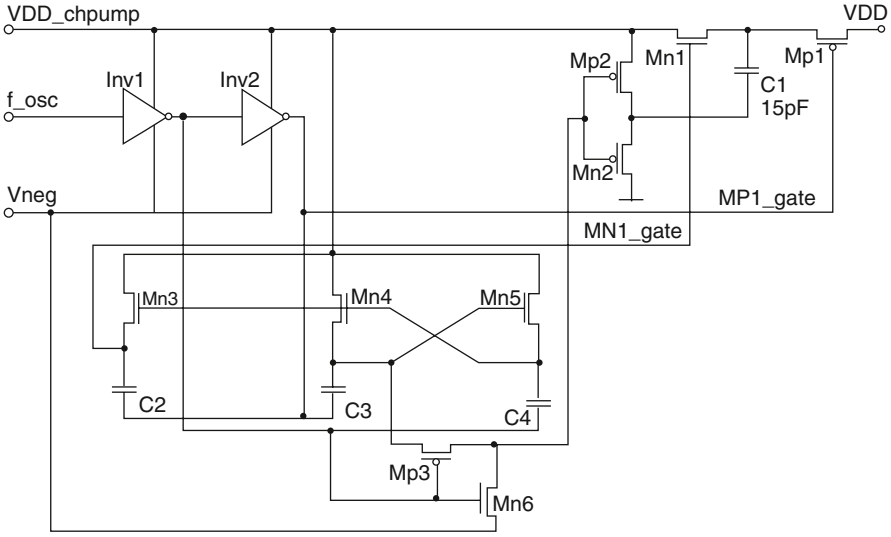


Fig. 12 Power charge pump as UHF DC/DC up converter

reduced to *VDD_chpump* potential, Mn1 gets non conductive, the bottom node of C1 is pushed to VDD level and Mp1 is motivated to conduct with a *Vneg* potential at its gate. The generation of the control signals consumes in average about 50 nA from the *VDD_chpump* potential of 600 mV. The average power consumption from the secondary charge pump *V_neg* is below 20 nA.

3 Local Oscillator

In the HF operation mode the local clock can be extracted from the 13.56 MHz carrier. The power consumption of such a HF clock generator is typically below 500 nW. At about 900 MHz the necessary power for such a field clock extraction would be bigger than for the rest of the chip. That is the reason why local oscillators are used for passive UHF transponder chips. UHF Protocols for contactless communication consider this certainty. The architecture of the oscillator in Fig. 13 shows the proposed regenerative first order oscillator [17, 18].

The most severe specification features of the local oscillator are an oscillator frequency larger than 1.92 MHz [19], a minimum operation voltage of 550 mV and a power consumption below 500 nW. In the transient timing of Fig. 14 the positive edge of *Vswitch* will start to toggle the inverter in the input stage.

Capacitor C1 is discharged until the voltage drop over the resistor R1 and the resulting current is smaller than the current delivered by the PMOS current source Mp2.

$$\Delta V_{ramp} = \frac{I_{Mp1/Mp2} \cdot t}{C_1} \approx I_{Mp2} \cdot R_1 \tag{13}$$

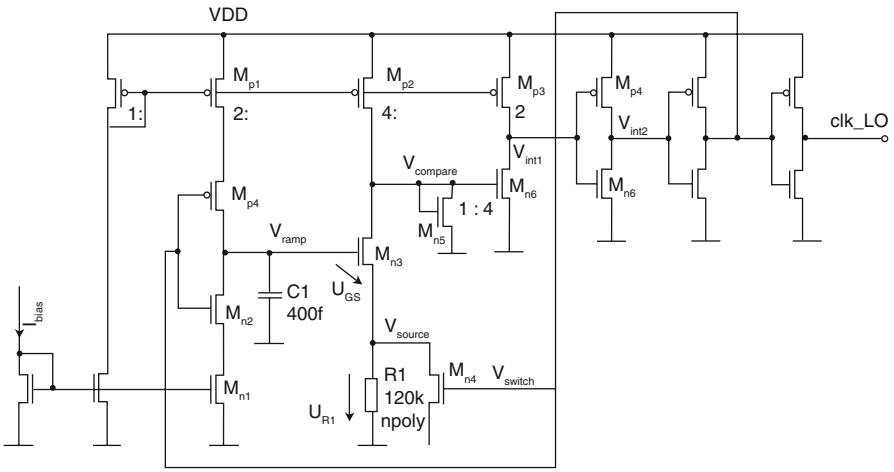


Fig. 13 Transistor level design of the local oscillator

In Formula 13 the integrator node *Vramp* is correlated with the maximum voltage over R1 where the oscillator will toggle in the second state. Is the current through R1 bigger than the constant current delivered by Mp2, the voltage potential at *Vcompare* will drop and the inverter Mn6/Mp3 will toggle. Diode Mn5 will clamp the voltage to one U_{GS} which reduces the voltage dependency of this circuit. The final oscillator period T in the first order only depends on values of passive components.

$$T \approx 2 \cdot n \cdot R_1 \cdot C_1 \tag{14}$$

In the second phase the NMOS switch Mn4 is bypassing R1 and in the first moment a big parasitic current will be conducted via Mn3 to VSS. *Vramp* is integrated down

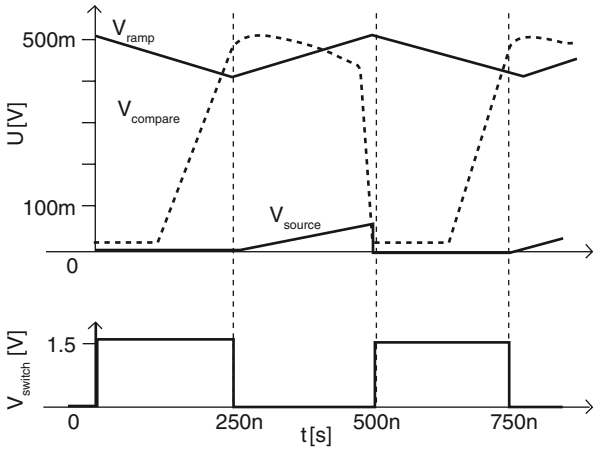


Fig. 14 Timing of some local oscillator nodes

until the U_{GS} of Mn3 is not able to sink the current delivered by Mp2 any more, node *Vcompare* will rise and the oscillator gets back again into the first state.

4 Units for Contactless Communication

Reader to transponder communication is called downlink or RxD. ASK is used in both EPC standards but the modulation index is different. A modulation index of 10–30% is specified for EPC HF and 100% ASK for EPC UHF. The consequence is the implementation of two optimized RxD units—a more sensitive and precise HF demodulator and a very low power UHF demodulator. A precondition for the following demodulator concepts is a constant chip input impedance. Under this constant impedance the coil voltage will then change proportional to the change of the field strength initiated by the reader.

4.1 HF RxD Unit

During a 10% index field strength gap the coil voltage will drop from 4 V_{peak} to about 3.4 V_{peak}. It should also be mentioned that the falling and rising edge steepness are limited because of the narrow bandwidth in the resonance circuit when a high quality factor Q is aspired. Additionally the chip impedance is not constant when the coil voltage is reduced and the typical consequence is a reduction of the modulation index.

The architecture [20] of the HF RxD unit is able to detect a minimum modulation index of 4% which can be adjusted well with the window generation module (Fig. 15). Two of the threes output signal of the window generator are fast enough

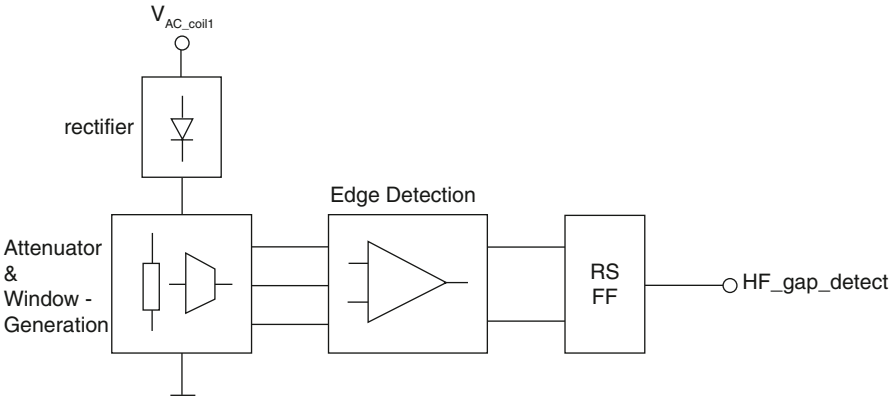


Fig. 15 HF window comparator architecture

to follow the subcarrier signal well. The voltage potential in between is defined by a slow integrator with a slop of 9 mV/µs which is slow compared to two outer signals with a minimum steepness of 33.6 mV/µs. Two comparators will detect the crossing of the slow inner signal with one of the fast outer signals and an attached flip flop will be set.

4.2 UHF RxD Unit

As the coil voltage is below 1 V_{peak} at incoming power levels of −6 dBm, a peak detection architecture [21] with a MOS rectifier diode cannot be used any more. Antenna voltage AC coupling, DC operation point biasing and an attached comparator are the important conceptual topics for the UHF RxD unit shown in Fig. 16.

A detailed transistor level schematic of the RxD unit is shown in Fig. 17.

C_{AC} is used to couple the antenna peak voltage information which has a dynamic range of 0.7–1.5 V to the internal node *offset*. Mn1 is a source follower which defines the lowest peak potential of node *offset* according to the gate potential U_{DCbias}. Transistors Mp2 and Mn1 are designed to define the minimum low potential at about 110 mV. To avoid during the positive coil voltage peaks which could cause an oxide-damage, diode Mp5 is used to clamp node *offset* to one diode voltage above VDD. The comparator is represented by transistor Mp1 and Mn3. Drops the voltage at node *offset* several periods below the threshold of Mp1, node *curr_comp* will also drop and the second stage of the comparator *amp_stage* will signal an UHF gap with a HIGH at the output node *UHF_gap_detected*. The gap between the threshold of transistor Mp1 and the 110 mV defined by Mn1 represents the small UHF coil voltage where the UHF Rxd unit detects a gap.

In Fig. 18 the start of a power gap with internal demodulator signals is shown. With the precondition of a constant chip impedance the coil voltage *VAC_coil1* drops proportional to the power gap which has been initiated by the reader. Because of parasitic capacitors, node *offset* in Fig. 18 oscillates with 868 MHz and with an amplitude reduced by 20%. The lower peak voltage potential of node *offset* is slightly below 90 mV (instead of 110 mV). When the high potential of signal offset is dropping below the threshold voltage of 450 mV of transistor Mp1, the low pass filtered comparator output signal *amp_stage* switches to high and informs the digital state machine about the beginning of a field gap.

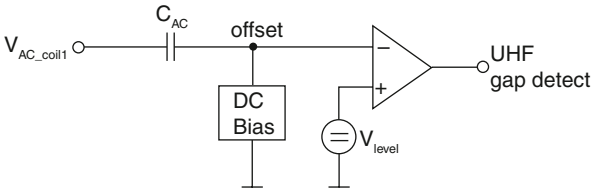


Fig. 16 UHF demodulator architecture

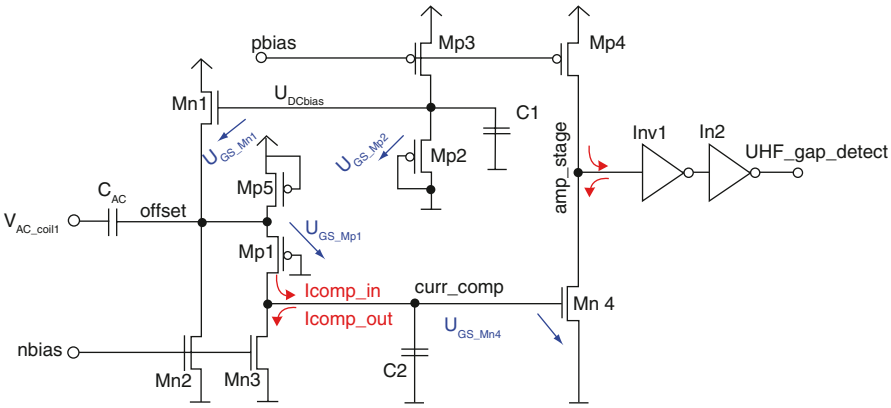


Fig. 17 UHF demodulator

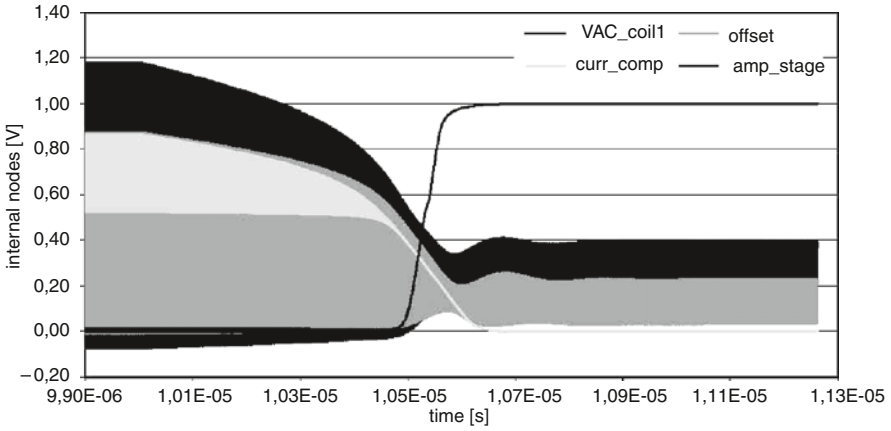


Fig. 18 UHF demodulator transient response when a UHF field gap occurs

4.3 Shunt, TxD—Load Modulator and Backscatter Unit

Equation 12 showed that only a small equivalent serial chip input impedance will guarantee a coil voltage big enough to power the chip. But in high power conditions without a parallel voltage stabilization especially in high magnetic field strength conditions the big coil voltage would damage the coil to chip interface. Parallel to the shunt stabilization unit the TxD communication unit is necessary to perform the load modulation and the backscatter operation. In Fig. 18 a single power transistor solution [22] is shown which includes a novel fast regulation loop for HF load mod-

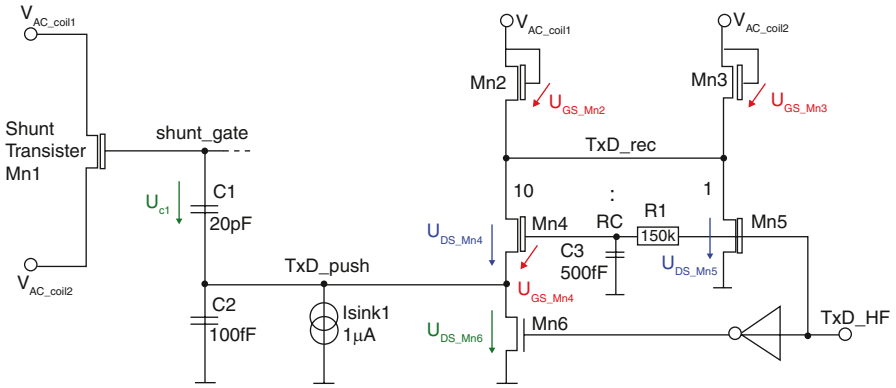


Fig. 19 Shunt and TxD unit

ulation which is untypical for typical passive RFID chips. With this single NMOS transistor solution the parasitic junction capacitance is reduced to a minimum. In all HF RFID standards a synchronous half duplex communication link is defined. To stay synchronous during the communication phase, it is necessary to be able to extract all the time the carrier clock. So a coil voltage of 700 mV_{peak} is necessary during load modulation.

At node *shunt_gate* a coil voltage regulation loop with two regulation speed states is connected. As regulator device an integrator is used with C1 as integrator output-capacitor. One speed state is the closed loop regulation and the other state is the loop opened and the regulation point is frozen in the voltage U_{C1} of capacitor C1. In the HF mode the coil voltage is regulated via the shunt transistor Mn1 to about 4 V_{peak}. As in Fig. 20 shown load modulation starts with the rising edge of node *TxD_HF*. Transistor Mn5 will pull slightly down the coil nodes *VAC_node1/2* and it will already start to discharge parasitic charge stored at node *TxD_rec*. After a delay of a few ns, Mn4 starts to conduct and charge will be transferred to the regulator bottom node *TxD_push*. As in Fig. 20 shown, an alternating voltage of a few hundred mV is enough to pull down the coil voltage. The charging phase stops, when UDS_Mn4 is close to VSS. This results in a stable coil voltage at:

$$V_{AC\_coil1/2\,peak} = U_{GS\_Mn2} + U_{DS\_Mn4} + \Delta U_{GS\_Mn1}$$

At a field strength of 1.5 A/m the peak voltage during load modulation is at pulled down to 1.2 V (see Fig. 20).

In the asynchronous UHF mode a PMOS transistor will pull node *TxD_push* to the VDD potential which causes a coil voltage close to zero and good backscatter information. The dynamic in the antenna voltage is less critical [23] compared to the load modulation in the HF band.

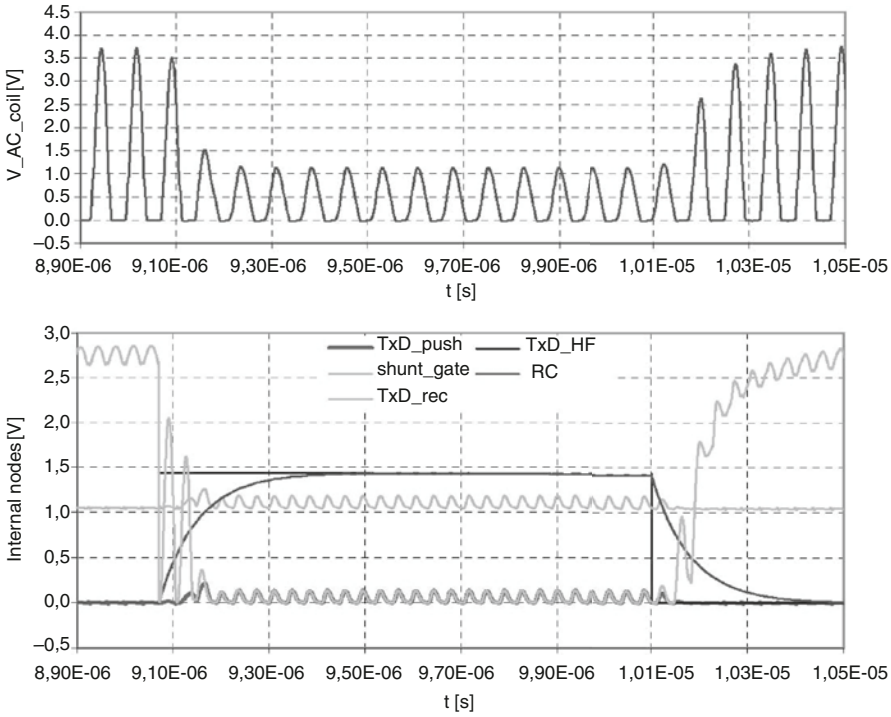


Fig. 20 Timing during HF load modulation

5 Conclusions

This work has shown an analog frontend which supports the EPCglobal HF and UHF standards. A balanced antenna to chip interface with one rectifier, two secondary charge pumps and two DC/DC converters have been presented and discussed. A compact low power and low voltage 2 MHz relaxation local oscillator with good power supply and bias rejection has been shown. Methods to build a compact UHF RxD unit and a well controlled TxD HF load modulator with only one shunt/modulator transistor are completing the frequency comprehensive analog front end.

References

1. K. Finkenzeller, RFID-Handbuch, Grundlagen und praktische Anwendung induktiver Funkanlagen, Transponder und kontaktloser Chipkarten, Hanser Fachbuchverlag, 4th edn. (2006). ISBN 3-446-40398-1
2. GS1 EPCglobal, EPC Radio-Frequency Identity Protocols Class-1 Generation-2 UHF RFID, Protocoll for Communications at 860–960 MHz Version 1.2.0, edition 1.2.0. (EPCglobal Inc., 2007)

3. GS1 EPCglobal, *EPC Radio-Frequency Identity Protocols HF Version 2*, edition 0.3. (EPCglobal Inc., 2007)
4. D. Engels, A Comparison of the Electronic Product Code Identification Scheme and the Internet Protocol Address Identification Scheme, (Technical memo: Auto-ID Center, Cambridge, 2002)
5. R. Herschmann, M. Camp, H. Eul, Design und Analyse elektrisch kleiner Antennen für den Einsatz in UHF RFID Transpondern. Adv. Radio Sci. **4**, 93–98 (2006)
6. P. Leuchtmann, *Einführung in die elektromagnetische Feldtheorie*, (Pearson Education, 2005), ISBN 3-8273-7144-9
7. R. Collin, S. Rothschild, Evaluation of antenna Q, IEEE Trans. Antenn. Prop. **12**(1), 23–27 (1964)
8. L. Mayer, A. Scholtz, A Dual-band HF/UHF Antenna for RFID Tags, *Vehicular Technology Conference, 2008*. VTC 2008-Fall. IEEE 68th, 21–24 September 2008, Calgary
9. J. Curty, M. Declercq, C. Dehollain, N. Joehl, Design and optimization of passive UHF RFID systems, 1st edn. (Springer, 2007), ISBN 0-387-35274-0
10. U. Karthaus, M. Fischer, Fully integrated passive UHF RFID transponder IC with 16.7 μW minimum RF input power, IEEE J. Solid-State Circuits, **38**(10), 1602–1608 (2003)
11. U. Friedric, U. Karthaus, S. Furic, H. Saeppa, The Palomar project, Funded by the European commission within the IST program D7, 2002
12. D.M. Dobkin, The RF in RFID—passive UHF RFID in practice, in *Newnes Communication Engineering Series*, 1st edn. (2007), ISBN 978-0750682091
13. E. Bergeret, J. Gaubert, P. Pannier, Standard CMOS voltage multipliers architectures for UHF RFID applications: study and implementation, in *IEEE International Conference on RFID*, 2007
14. A. Facen, A. Boni, A CMOS analog frontend for a passive UHF RFID tag, low power electronics and design, in *ISLPED'06. Proceedings of the 2006 International Symposium on Low power electronics and design*, 2006
15. A. Missoni, C. Klapf, W. Pribyl. G. Hofer G. Holweg, A triple-band passive RFID Tag, in *ISSCC* 2008, Session 15.2, San Francisco, USA
16. T. Umeda, H. Yoshida, A 950-MHz rectifier circuit for sensor network tags with 10-m distance. IEEE J Solid State Circuits. **41**(1), 35–41 (2006)
17. C. Klapf, A. Missoni, W. Pribyl, G. Hofer, G. Holweg, Analyses and design of low power clock generators for RFID TAGs, in *IEEE PRIME* 2008, Istanbul
18. S.L.J. Gierkink, Control linearity and jitter of relaxation oscillators, in *Proefschrift (Dissertation)*, ISBN: 90-36513170
19. C. Diorio, Watching the clock—a tag's clock frequency is critically important to tag performance. RFID J., (2006)
20. A. Missoni, Circuit arrangement for the analog-digital conversion of a voltage (ASK-) modulated by modifying the amplitude between a low level and a high level WIPO international patent, Pub.Nr.: WO/2004/036860, 2004
21. Z. Zhu, B. Jamali, P. Cole, An HF/UHF RFID analogue front-end design and analysis, Autoidlabs-WP-HARDWARE-012, Sept. 2005, white paper series/edition 1
22. Magellan, Shunt Regulator, International Patent WO89/07295, 1989
23. P. Nikitin, K. Rao, in *IEEE Antennas and Propagation Society International Symposium 2006*, 9–14 July 2006, pp. 1011–1014

Printed Electronics—First Circuits, Products, and Roadmap

Jürgen Krumm and Wolfgang Clemens

1 Introduction

With the advent of soluble semiconductors and conductors, the field of printed electronics became possible. Printed electronics is often associated with organic electronics where the characteristic feature is the use of organic materials in essential parts of the devices and circuits. Both fields overlap in many aspects. In printed electronics, however, circuits do not rely on organic materials but are manufactured by means of low-cost, large-area, and high-volume printing methods [1]. In consequence, electronic products fabricated in large quantities like newspapers on a daily basis give rise to novel applications for integrated electronics like radio-frequency identification (RFID) labels replacing optical barcodes, intelligent packaging and smart objects for processing and displaying information, and flexible displays.

The major advantage of printed electronics is its ease of fabrication rather than improved electrical performance of printed circuits. A large number of organic and inorganic materials can be deposited from solution, allowing e.g. application of high-speed and high-throughput printing methods. With increased fabrication speed and output, lower fabrication costs compared to conventional fabrication technologies are possible.

Currently, low-cost printed electronics is still in the early stages of development and optimization of devices and processes. A variety of materials and fabrication processes are explored and optimized. One major issue in printed electronics is the development of electronic inks which can be used in the printing process while still maintaining their semiconducting or conducting behavior. Moreover, products fabricated with printed electronics are still in the early stages. This article presents an overview over the field of printed electronics, recent developments in circuit design with focus on RFID circuits, first products and future directions of the development.

J. Krumm (✉)
PolyIC GmbH & Co. KG, Tucherstraße 2, 90763 Fürth, Germany

H. Casier et al. (eds.), *Analog Circuit Design,*
DOI 10.1007/978-94-007-0391-9_17, © Springer Science+Business Media B.V. 2011

2 Printed Electronics

In printed electronics, printing techniques like flexography, gravure printing, screen printing, or ink jet printing, etc. are readily available. Printing of electronic circuits considerably increases fabrication volume and throughput as well as considerably cuts fabrication costs. These advantages are specific to printed electronics. One key concept of printed electronics is roll-to-roll manufacturing, where a printing substrate is fed through a printing machine from one roller to another.

A popular example of roll-to-roll fabrication is gravure printing whose operation principle is illustrated in Fig. 1. In gravure printing, a flexible printing foil is continuously unwound from a roll, travels between a rotating gravure cylinder and an impression cylinder, and is finally rewound onto another roll. The rotating gravure cylinder is wetted by an ink in a fountain. The liquid is scraped by a doctor blade so that only the quantity filling up the recessed cells engraved into the cylinder remains. The remaining ink is then transferred to the target position on the printing substrate while the cylinder rolls across it.

In conventional printing, the ink usually colorizes the printing substrate so to yield a visual impression. In printed electronics, "electronic inks" are used in order to apply functional layers of the electronic circuits (electrode, semiconductor, and insulator) in successive printing steps.

A typical layer setup for printed circuits [3] is shown in Fig. 2. The depicted setup consists of the substrate and four functional layers. These functional layers are the lower electrode, the semiconductor, the insulator, and the upper electrode.

The substrate carries the subsequently deposited layers throughout the manufacturing process. It usually has no electrical function in the circuits. It is, however, required to provide reasonable mechanical flexibility for the manufacturing process in

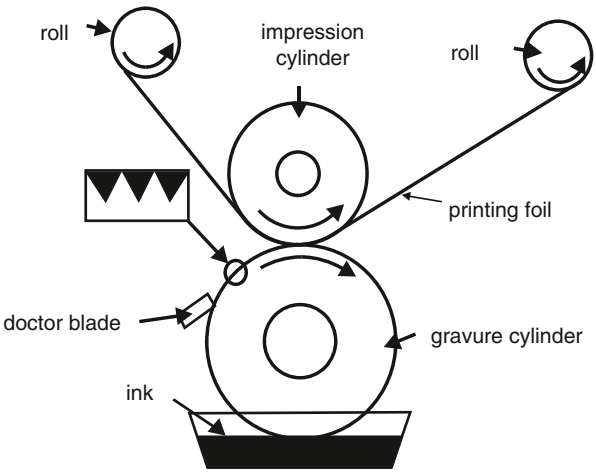


Fig. 1 Operating principle of gravure printing [2]

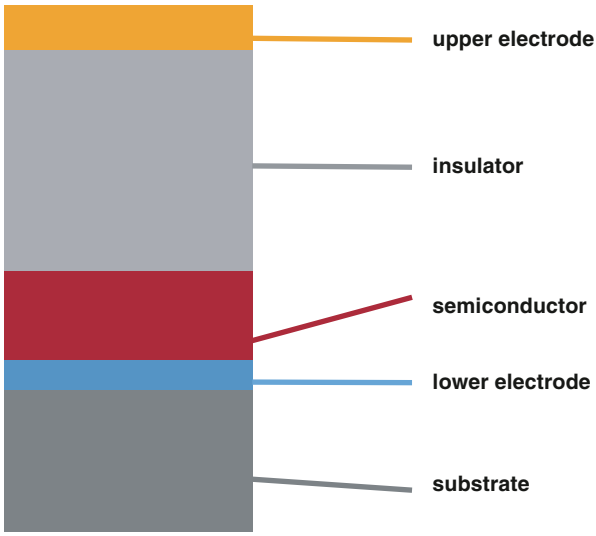


Fig. 2 Layer setup of a typical printing process [3]

use. Materials like polyethylene terephthalate (PET) can be used. The four functional layers are successively stacked on top of the substrate in order to implement integrated circuits.

The first functional layer is the lower electrode layer. It is used as a first interconnect layer and for electrodes of semiconductor devices like transistors, diodes, or capacitors. Electrode materials can be metals like silver, copper [3] or metal-like organic materials such as PEDOT-PSS [4].

Stacked on top of the lower electrode is the semiconductor, which is typically an organic semiconductor like poly(3-alkylthiophene) (P3AT) [3], polytriarylamine (PTAA) [4], polyfluorene (F8T2) [5] or an inorganic semiconductor such as carbon nanotubes [6].

The next layer following the semiconductor is the insulator layer. The insulator separates the semiconductor and the upper electrode.

The top most layer is the upper electrode. Like the lower electrode, it is used for wiring and as electrode material for devices such as transistors, diodes, or capacitors.

The four-layer setup in Fig. 2 provides all types of devices used in basic integrated circuits. In most circuits, transistors are the main building blocks. Thin-film transistors (TFTs) are usually applied in printed electronics. These devices consist of a source/drain layer (lower electrode in the four-electrode setup), a thin film of semiconducting material, an insulator, and a gate electrode (upper electrode layer in the four-layer setup).

Figure 3a illustrates the general design of a printed TFT realized with the four-layer setup from Fig. 2. In the figure, G denotes the gate electrode in the upper electrode layer and S/D the source/drain electrodes in the lower electrode layer. The output characteristics of printed TFTs are comparable to conventional MOSFETs (see Fig. 3b for a typical example using a p-type P3HT transistor) in shape but transistor parameters such as threshold voltage or charge-carrier mobility are generally

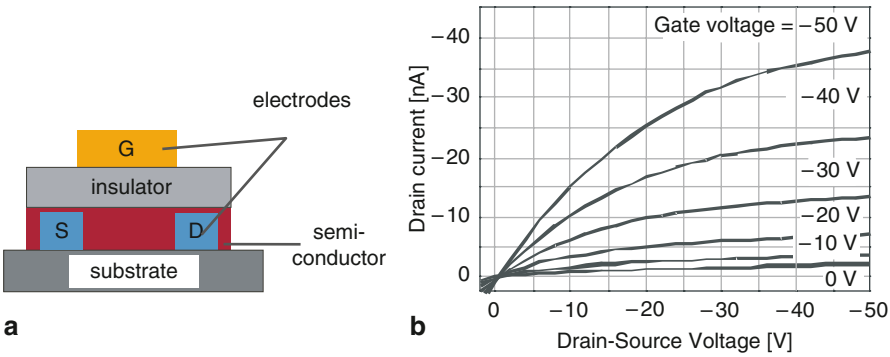


Fig. 3 Printed thin-film transistor: **a** structure, and **b** output characteristics [20]

less convenient for chip designers than those of conventional silicon transistors. For example, charge carrier mobilities of printed TFTs are typically below 0.1 cm²/Vs as opposed to 500 cm²/Vs and above for conventional silicon. Moreover, the device parameters can considerably differ between different types of printed TFTs owing to their individual choices of materials and fabrication technologies.

In order to implement devices like printed TFTs the fabrication process has to meet certain requirements. For example, continuous line structures are required for the drain and source electrodes with attainable feature sizes in the micron range (≥15 µm [3]) so to provide reasonable performance figures for the circuits. The semiconductor and the insulator have to be stacked upon each other as thin, homogeneous, and defect-free films. No fabrication step is allowed to harm the electrical performance of previously deposited materials beyond proper use. Moreover, the upper electrode must be deposited with precise registration, i.e. accurate alignment relative to the structures on the underlying layers. All materials used in the process have to be compatible with the manufacturing steps in which they are handled.

3 First Printed Circuits

By using soluble materials, prototype circuits with complexities of up to 2,000 transistors have been demonstrated [7]. Milestone circuits with focus on RFID applications are 64/128-bit RFID transponders with pentacene [7, 8] and a 64-bit transponder with poly(3-hexylthiophene) [9, 10] as semiconductor materials. These circuits were manufactured using clean-room techniques. Owing to their low-cost and high-throughput fabrication technologies, printed circuits currently lag behind these complexities and also in performance figures like switching speed or packing density. In the domain of printed electronics, first completely printed RFID transponders were realized in 2007 [11]. These transponders provided an oscillation signal but no specific ID code. State-of-the-art results on complex logic circuits are 4-bit code generators for RFID transponders as demonstrated by [6] and [3]. The circuit in [6] was realized using gravure printed transistors with carbon nanotubes as semiconductor

material but was clocked with 1 Hz by an external function generator. The printed circuit in [3] contains all blocks necessary for a Manchester-encoded RFID transponder (rectifier, clock generator, data generator, modulation transistor). Therefore, it will be examined subsequently. In order to give an impression on the switching speeds currently reached with printed transistors, a printed ring oscillator discussed in [3] is briefly shown. Then, results on the code generator chip are presented.

3.1 Printed Ring Oscillator

A basic type of circuit based on printed transistors is the ring oscillator. This circuit consists of an odd number of inverters in a chain with the output of each inverter connected to the input of the following. The output of the final inverter is fed back to the input of the first. Ring oscillators are often used as benchmark circuits in order to determine the switching speeds and logic capabilities of semiconductor technologies. Employing a ring oscillator is a simple test to analyze the feasibility of transistors and their geometrical design for logic circuits. As long as the ring oscillator produces an oscillation the chain of inverters is capable of regenerating input signals. This regeneration property is an important prerequisite for logic circuits [12, 13].

In [3], a printed ring oscillator with 15 inverters in a chain is presented. The circuit starts oscillating at a supply voltage of 6 V and reaches an oscillation frequency of 107 Hz at a supply voltage of 15 V. Figure 4 details the schematic and the

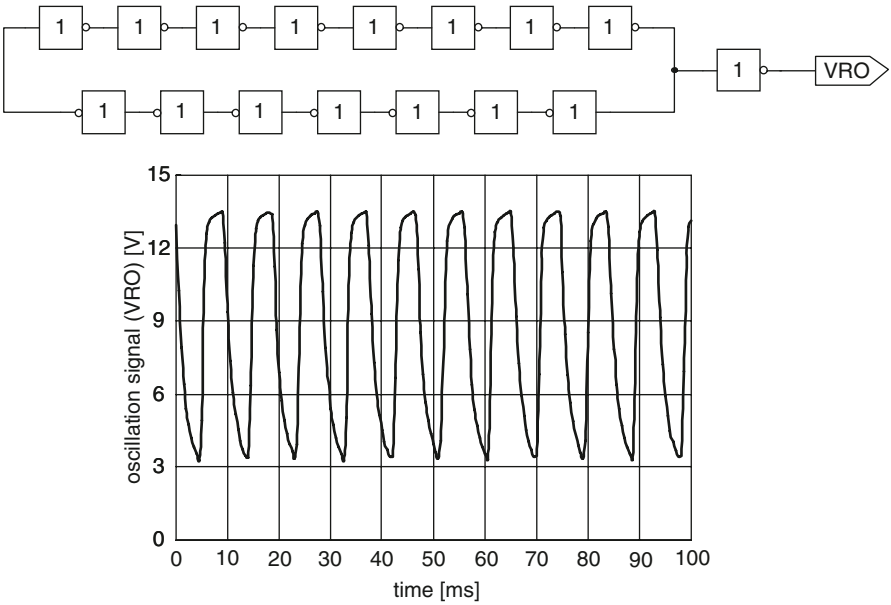


Fig. 4 Schematic and measured oscillation signal of a 15-stage printed ring oscillator at an operation frequency of 15 V [3]

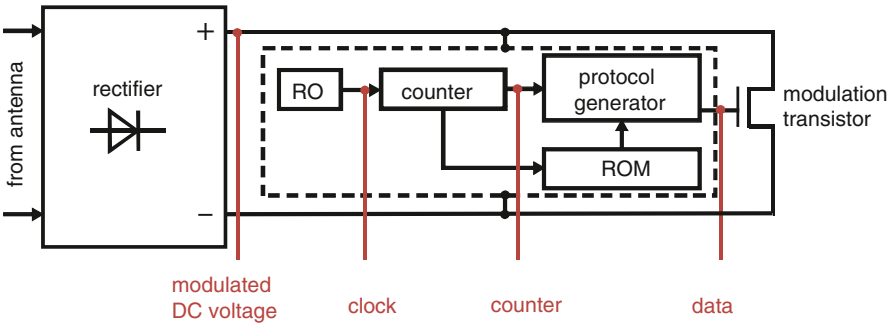


Fig. 5 Schematic block diagram of a printed 4-bit code generator chip. (Adaption from [3])

measured oscillation signal at a supply voltage of 15 V. The transistors are based on P3HT as semiconductor material.

3.2 Printed 4-Bit Manchester Chip

While ring oscillators are useful in testing the performance of transistors, more sophisticated circuits with numerous building blocks are needed for reasonable application circuits. One such example is a printed 4-bit code generator chip for RFID applications presented in [3]. This chip consists of numerous building blocks: a rectifier, a ring oscillator with 15 stages as clock generator, a counter with 3 flip-flops, a protocol generator for Manchester-encoded data signals, and a read-only memory. Figure 5 shows the block diagram of the code generator chip, which consists of about 200 individual devices.

The chip is designed for operation in a passive transponder by drawing energy from the radio-frequency field of an RFID reader. In the rectifier, the induced AC voltage with a frequency of 13.56 MHz at the transponder antenna is rectified so to generate the DC supply voltage for the remaining components of the chip. The logic components in Fig. 5 generate a data stream with Manchester encoding. This data stream consists of a start sequence with idle time and start bit and a data sequence of 4 bits. The RFID reader is synchronized with the clock rate and phase of the transponder by use of this start sequence. The data sequence is fed to the gate of a modulation transistor. When this modulation transistor is switched on (on-state), the transponder circuit draws a certain amount of energy from the radio-frequency field of the reader. In the off-state, when the modulation transistor is switched off, the transponder circuit only draws a fraction of the energy consumed in the on-state. This variation in energy consumption is used to transmit information to the reader.

Figure 6 shows the measured clock signal, the output of the final (3rd) flip-flop, the data sequence of the code generator, and the voltage at the output of the rectifier. The chip was programmed to issue a Manchester-encoded data sequence of 1-0-0-1. It was measured with a function generator attached to the input of the rectifier. In

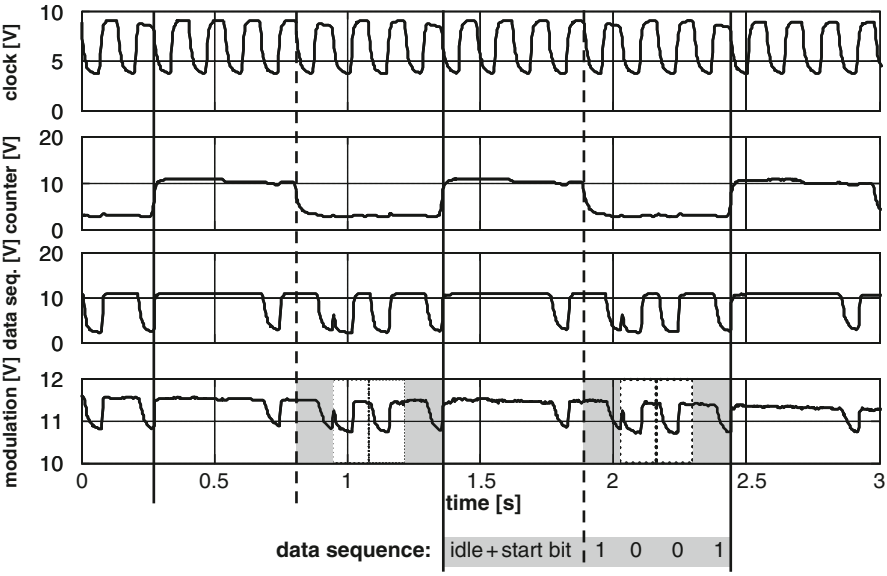


Fig. 6 Measured voltages in the 4-bit code generator. (Data from [3])

this measurement setup, the modulation caused by the modulation transistor was detected by monitoring the variation of the DC voltage at the output of the rectifier. In the off-state, a certain DC voltage is supplied by the rectifier. In the on-state, when the modulation transistor sinks more current, this supply voltage drops by a certain amount. The variation in the DC supply voltage is monitored in order to detect the modulation.

4 First Products and Roadmap of Printed Electronics

Due to the availability of printable electronic materials, different products based on printed electronics are possible. Examples of these are: intelligent packaging, flexible displays, wearable electronics in garments or printed RFID transponders for EPC (electronic product code) applications. The majority of these applications are still in the early stages and require coordinated work regarding e.g. materials or fabrication technologies. In the field of organic and printed electronics, a roadmap was compiled which discusses the major applications and their requirements. In the following, this roadmap will be discussed.

4.1 Roadmap for Organic Electronics

In standard semiconductor technology, the “International Technology Roadmap for Semiconductors” (ITRS, [14]) has been compiled and is regularly updated in

order to create a common understanding of the future development of the industry and to coordinate the activities of the involved companies and research institutes. In the domain of organic and printed electronics, a similar roadmap exists with the title "OE-A Roadmap for Organic and Printed Electronics" [1]. This roadmap was initiated and is maintained by the Organic Electronics Association (OE-A, [15]), a working group within the German engineering federation VDMA [16]. The OE-A was founded in 2004 and has more than 120 members [16] (figure for 2009). These members are from industry and scientific institutions in Europe, USA, Canada, Australia, Taiwan, South Korea, and Japan. The OE-A supports all activities in organic electronics, specifically development in the following domains:

- materials (substrate, semiconductors, conductors, etc.),
- technology (printing, machines, fabrication expertise),
- device (transistors, circuits, organic photovoltaic devices, etc.),
- products, applications, markets.

In their roadmap, the OE-A members identified nine major application fields of organic and printed electronics that are expected to have the highest market potentials. These application fields are:

- organic photovoltaics,
- flexible displays,
- electroluminescent and OLED (organic light emitting diode) lighting,
- printed RFID,
- printed memory,
- organic sensors,
- flexible batteries,
- smart objects,
- smart textiles.

For each of these fields, the OE-A has made individual subroadmaps in the framework of [1]. These application-specific subroadmaps define estimated performance potentials of commercial products with respect to the short-term, medium-term, and long-term future. Additionally, key performance parameters are discussed as well as critical challenges.

All fields have specific advantages and limitations with respect to existing solutions based on conventional electronics. Major advantages of organic and printed electronics in nearly all application fields are, however, their low thickness and high flexibility, the potential for high volume production at low cost and consequently the potential to enable electronics in areas where today's conventional electronics will hardly be integrated. However, there are some limitations, too. These are due to the typically weaker electronic properties of the materials (e.g. conductivity, charge carrier mobility, life time) and the processes (printing vs. clean room processes). Therefore, organic and printed electronics is in most cases not expected to substitute standard electronics but will enable new application areas, where the properties of printed electronics give significant advantages.

Table 1 Performance evolution of major application fields over time [1]

Application field	Estimated availability/evolution		
	Short-term (2009–2012)	Medium-term (2012–2017)	Long-term (2018+)
Organic photovoltaics	Consumer and first off-grid applications	Off-grid power, building integration	Grid-connected power generation
Flexible display	Price labels, e-readers	High resolution color e-readers, e-posters	Electronic wall papers, rollable OLED TVs
OLED/EL lighting	Small lamps, design and decorative applications	Light tiles, technical/architectural lighting	Flexible lighting elements
Printed RFID	Brand protection, e-ticketing	Logistics and automation	Item-level tagging, EPC, identification
Printed memory	Brand protection, identification, games	High end brand protection, advanced games	Electronics, multimedia
Organic sensor	Photodiode, temperature, pressure, chemical	Potentiometric sensor array	Intelligent sensor, embedded system
Flexible battery	Low capacity, discontinuous use	Higher capacity, continuous use	Direct integration into packages, systems
Smart objects	Greeting cards, animated logos	Intelligent tickets, initial smart packaging	Complex smart packaging
Smart textiles	Clothing integrated keypads, sensors, light effects	Clothing integrated display, photovoltaics	Fuel cells, fiber with integrated sensors

The anticipated evolution of the performance potentials in the nine application fields is briefly summarized in Table 1. Currently, only few products are commercialized in niche markets, e.g. applications with printed organic photovoltaics [17, 18]. Many products are expected to establish in the near future and to bring significant revenues in the medium and long term.

The reader is referred to [1] for a more detailed discussion of the individual items in the table. Here, only the subroadmap for printed RFID is further detailed by aid of Fig. 7.

4.2 Roadmap for Printed RFID

For printed RFID, market entry is anticipated to take place in the short-term future with basic products like brand protection and electronic ticketing. In these applications, tag complexities of 1–8 bits in the HF frequency range are implemented. As the performance evolves over time, applications in logistics and automation with tag complexities of 16–64 bits are targeted. As has been shown in Sect. 3, prototype circuits for all complexity levels have been demonstrated with organic electronics. The more complex prototypes were however fabricated using clean-room facilities. Therefore, general availability of tag complexities of 96 bits in the HF and UHF

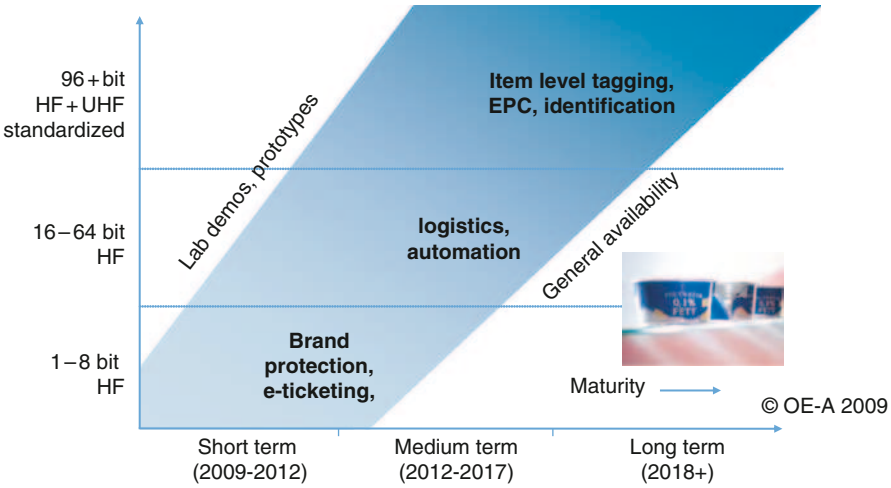


Fig. 7 Evolution of performance potential of printed RFID versus time [1]. (Used by courtesy of OE-A)

frequency range for real printed products with item level tagging and product identification is expected in the long-term future of 2018 and beyond.

The subroadmap for printed RFID identifies some key parameters on the application level. These parameters define the potential applications of printed RFID transponders. While brand protection and electronic ticketing can be realized with basic requirements, applications like logistics and automation depend on more mature performance figures for the following parameters:

- number of ID bits,
- reading distance between tag and reader,
- reading frequency, which defines the possibility of UHF systems,
- cost per tag,
- bulk reading: anti-collision detection of multiple tags in the reading range of the reader.

It is straightforward to translate the key parameters on the application level into technology-related parameters on the device level. The resulting key parameters named by the subroadmap for printed RFID are:

- charge-carrier mobility of the semiconductor: defines attainable circuit speed,
- switching speed: one factor defining the read-out time for the ID bits; this parameter is affected by the charge-carrier mobility and printing resolution,
- operation frequency of the rectifier: defines the reading frequency, i.e. the availability of HF and UHF systems,
- printing resolution: defines the circuit yield, i.e. the cost per tag and the switching speed,

- number of transistors: defines the circuit complexity, i.e. the number of ID bits and availability of more complex concepts like bulk reading, as well as cost per tag (through yield).

According to the OE-A roadmap, critical issues with printed RFID (called red brick walls in the roadmap) are:

- fabrication yield versus transponder complexity: More complex circuits require higher functional yield.
- availability of CMOS processes: Currently, integration of n-type transistors together with p-type transistors in printed circuits is difficult. However, CMOS processes are desirable because of increased robustness of circuits and reduced power consumption.
- writable memories: The availability of programmable memories (one to many write cycles) is desirable for applications like logistics and would therefore promote the market share of printed RFID.
- UHF frequency range: In standard RFID applications, the UHF frequency range is popularly used in order to reach long reading distances in comparison to HF transponder systems. Availability of printed UHF transponder would again promote the market share of printed RFID.

4.3 First RFID Products

First prototype circuits already appeared in 2007/2008 with reduced complexities in comparison to conventional electronics. For example, radio-frequency transponders with printed organic chips were successfully tested by the Printed Smart Labels project (see [19]) in a field trial at two conferences. A total of about 400 transponders were issued in September 2007 at the Organic Electronics Conference and

Fig. 8 Radio-frequency transponder with printed organic chip [20] used in a field trial at the MEDIA-TECH Expo conference 2008

Exhibition (OEC07, Frankfurt, Germany) and 4,000 transponders at the MEDIA-TECH Expo conference in May 2008 (Frankfurt, Germany). These transponders consisted of printed polymer chips and conventional antennas [20]. They operated at a reading frequency of 13.56 MHz but did not generate individual ID codes. The tags were used for entrance control to the conference area and for counting of visitors. Figure 8 shows one of the fully-assembled transponders.

In order to enhance their functionality, printed RFID transponders can be equipped with interactive optical labels or sensors for tracking environmental conditions like temperature or humidity, or the presence of certain chemicals. Other components like memory, loudspeaker or battery packs can be added, too, in order to create printed "smart objects". The idea behind smart objects is to combine multiple electronic devices into applications ranging from simple animated logos to intelligent sensors [16].

5 Summary

Printed electronics is a novel platform technology enabling many application fields of low-cost, thin and flexible, as well as high-volume electronics. The most prominent applications are organic photovoltaics, flat displays, and RFID. The technology is still in an early stage and there are several limitations due to unfavorable material and process parameters. First products are currently entering the market in special applications. High-volume products are expected to be available in the next few years. Printed electronics is not regarded to substitute conventional electronics but gives rise to novel applications where tight requirements on fabrication cost or form factors cannot be met by traditional solutions.

References

1. Organic Electronics Association, OE—a roadmap for organic and printed electronics (Dec, 2009), available online at http://www.vdma.org/wps/portal/Home/en/Branchen/O/OEA/OEA_News_StK_20091211_WhitePaper?WCM_GLOBAL_CONTEXT=/wps/wcm/connect/vdma/Home/en/Branchen/O/OEA/OEA_News_StK_20091211_WhitePaper. Accessed 9 Feb 2010
2. J. Krumm, Circuit analysis methodology for organic transistors, Dissertation, University of Erlangen-Nuremberg, Germany, 2008
3. W. Fix, R2R Printed Electronics. Presentation at Lope-C Conference, Frankfurt, Germany, June 23–25 (2009)
4. D. Zielke, A.C. Hübler, U. Hahn, N. Brandt, M. Bartzsch, U. Fügmann, T. Fischer, J. Veres, S. Ogier, Polymer-based organic field-effect transistor using offset printed source/drain structures. App. Phys. Lett. **87**, 123508 (2005)
5. G. Klink, E. Hammerl, A. Drost, D. Hemmetzberger, K. Bock, Reel-to-reel fabrication of integrated circuits based on soluble polymer semiconductor. Presentation at IEEE Polytronic (2005)

6. M. Jung, K. Jung, S.-Y. Lim, K. Lee, D.-A. Kim, J. Kim, J.M. Tour, G. Cho, All R2R Printable 4-Bits Digital Signal Processor for Printed RFID Using SWNT-TFTs, ICFPE, Jeju Island, Korea, Nov 11–13 (2009)
7. E. Cantatore, T. Geuns, A. Gruijthuijsen, G.H. Gelinck, S. Drews, D.M. de Leeuw, A 13.56 MHz RFID system based on organic transponders. IEEE ISSCC, 272, 273 (2006)
8. K. Myny, M.J. Beenhakkers, N.A.J.M. van Aerle, G.H. Gelinck, J. Genoe, W. Dehaene, P. Heremans, A 128b organic RFID transponder chip, including Manchester encoding and ALOHA anti-collision protocol, operating with a data rate of 1529b/s. IEEE ISSCC, 206, 207 (2009)
9. A. Ullmann, M. Böhm, J. Krumm, W. Fix, Polymer Multi-Bit RFID Transponder, ICOE Conference, Eindhoven, The Netherlands, 2007
10. W. Fix, A. Ullmann, R. Blache, K. Schmidt, Organic Transistors as Basis for Printed Electronics, in Physical and Chemical Aspects of Organic Electronics, ed. by C. Wöll, (Wiley-VCH, Weinheim, 2009), pp. 3–13
11. W. Fix, Polymer based 13 MHz RFID transponders, Organic Electronics Conference, Frankfurt, Germany, Sept 24–26 (2007)
12. C. F. Hill, Noise margin and noise immunity in logic circuits. Microelectronics. **1**, 16–21, April (1968)
13. S. De Vusser, J. Genoe, P. Heremans, Influence of transistor parameters on the noise margin of organic digital circuits. IEEE Trans. Electron Devices. **53**(4), 601–610, April (2006)
14. Homepage of International Technology Roadmap for Semiconductors, www.itrs.net
15. Homepage of Organic Electronics Association, www.oe-a.org
16. Organic Electronics Association, Organic and Printed Electronics, 3rd edn, (VDMA Verlag, Frankfurt, 2009)
17. E.C. Zeira, PV as a Market entry for printed electronics, presentation at the OEA Meeting, San Jose, USA, Dec (2008)
18. Konarka Technologies, Inc., Konarka opens world's largest roll-to-roll thin film solar manufacturing facility with one gigawatt nameplate capacity. Press release (Nov 24 2009), available online at http://www.konarka.com/index.php/site/pressreleasedetail/konarka_opens_worlds_largest_roll_to_roll_thin_film_solar_manufacturing_fac. Accessed 12 Feb 2010
19. Homepage of Printed Smart Labels (PRISMA) project, www.prisma-projekt.de
20. J. Krumm, Printed Electronics—From Vision to First Products, RFID Systech, 4th European Workshop on RFID Systems and Technologies, Freiburg, Germany, June 10, 11 (2008)

Towards EPC-Compatible Organic RFID Tags

Kris Myny, Soeren Steudel, Peter Vicca, Steve Smout, Monique J. Beenhakkers, Nick A. J. M. van Aerle, François Furthner, Bas van der Putten, Ashutosh K. Tripathi, Gerwin H. Gelinck, Jan Genoe, Wim Dehaene and Paul Heremans

1 Introduction

There has been significant interest in low temperature thin-film transistors (TFT) on foil in the last decade. The target applications are mainly backplanes for flexible displays and passive radio-frequency identification tags (RFID). The interest in flexible display lays in the user-friendliness of having a flexible, lower weight and less breakable medium whereas the interest in radio-frequency identification tag lies in the possibility to realize simple tags with significantly lower cost than traditional tags using Si-chips. This would allow a significant step forward in the realization of the broad vision of the "net of things", which has been slow to develop mostly because of economic reasons.

There are several thin-film technologies that are competing in this area. On the one hand we find established thin-film technologies like amorphous silicon or polysilicon where major effort has been invested to lower the process temperature or otherwise to develop transfer-technologies. On the other hand novel semiconductor materials have been introduced like metal-oxides, organic semiconductors, carbon nanotubes or chalcogenides. One that received considerable attention is based on conjugated organic molecules. Those organic semiconductors reach a similar mobility as amorphous silicon but are processed near room temperature. They allow the use of a broad range of deposition and patterning techniques ranging from conventional plate-based manufacturing, including photolithography, to novel deposition techniques, like printing.

In this chapter we will describe our progress regarding realizing complex chips (e.g. transponder chips) and complete RFID tags with organic semiconductors.

K. Myny (✉)
IMEC, Leuven, Belgium

K. Myny
Katholieke Hogeschool Limburg, Diepenbeek, Belgium

K. Myny
Katholieke Universiteit Leuven, Leuven, Belgium

H. Casier et al. (eds.), *Analog Circuit Design,*
DOI 10.1007/978-94-007-0391-9_18, © Springer Science+Business Media B.V. 2011

2 Organic RFID Tags

In recent years, several research groups have studied and published research results on organic RFID systems. In 2007, Cantatore et al. published a capacitively-coupled RFID system where a 64-bit code was read out at a base carrier frequency of 125 kHz [1]. The 64-bit code generator was fully functional at a 30 V supply voltage. In that pioneering work, lower bit generators (up to 6 bit) could be read out using a base carrier frequency of 13.56 MHz by a capacitive antenna. Ullmann et al. demonstrated a 64-bit tag working at a bit rate exceeding 100 b/s, readout by inductive coupling at a base carrier frequency of 13.56 MHz [2]. In this chapter, we review recent advances in both the digital transponder chip and the analog front-end of organic RFID tags, and demonstrate that organic electronics can result in a tag with a realistic code size, bit rate and reading distance at reasonable and allowed field strength [3–5].

The basic schematic of the organic RFID tag presented here is depicted in Fig. 1. The organic RFID tag consists of four different modules: the antenna coil, the HF-capacitor, the rectifier and the transponder chip with an integrated load modulator. The coil and the HF-capacitor form an LC tank resonating at the HF resonance frequency of 13.56 MHz, which provides the energy for the organic rectifier with an AC voltage at 13.56 MHz. The rectifier generates the DC supply voltage for the 64-bit organic transponder chip, which drives the modulation transistor between the on- and off-state with a 64-bit code sequence. Load modulation can be obtained in two different modes, depending on the position of the load modulation transistor in the RFID circuit, shown in Fig. 1. AC load modulation, whereby the modulation transistor is placed in front of the rectifier, sets demanding requirements to the OTFT, since it has to be able to operate at HF frequency. This is not obvious, as a consequence of the limited charge carrier mobility of the OTFT, being 0.1–1 cm^2/V s for pentacene as organic semiconductor. Therefore, load modulation at the output of the rectifier (DC load modulation) is preferred in organic RFID tags. In latter mode, the OTFT does not require to operate at HF frequency. The organic RFID tags in this chapter operate in DC load modulation mode. Nevertheless, organic RFID tags operating in AC load modulation mode have also been achieved [4].

2.1 Technology

In this section we describe the technology used to create high-performance organic RFID tags [4]. As mentioned earlier, the tags are composed of four flexible foils, with the following components: an inductor coil, a capacitor, a rectifier and a transponder. The coil is made from etched copper on foil, manufactured by Hueck Folien GmbH.

The HF-capacitor consists of a metal-insulator-metal (MIM) stack, processed on a 200 μm thick flexible polyethylene naphthalate (PEN) foil (Teonex Q65A, Dupont Teijin Films). The insulator material used for the capacitor is Parylene diX SR.

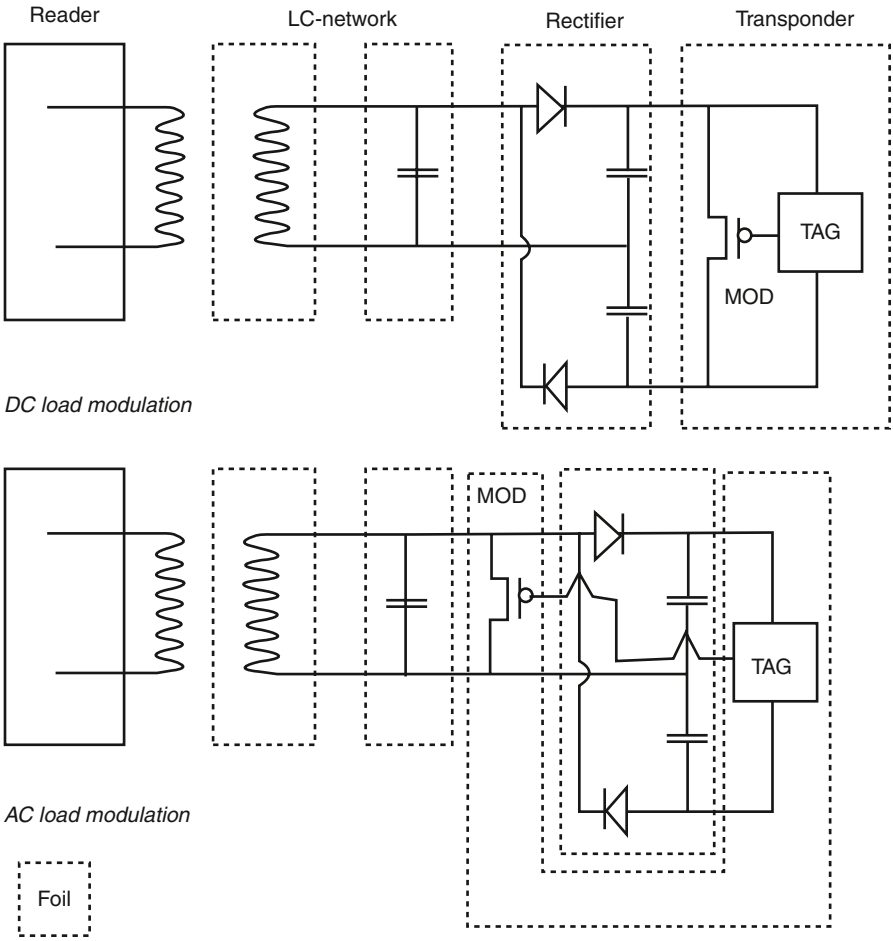


Fig. 1 Inductively-coupled organic RFID tags using DC (*top*) and AC (*bottom*) load modulation

The rectifier comprises two vertical Schottky diodes, and two capacitors in a so-called double half-wave configuration [3]. The schematic of the rectifier is shown in Fig. 7, and a photograph of the rectifier is depicted in Fig. 2. The substrate for manufacturing the rectifiers is a 200 µm thick, flexible 150 mm PEN foil, on which first a metal-insulator-metal (MIM) stack is processed for the capacitors in the circuit. The metal layers are 30 nm of gold (Au) and the insulator is Parylene diX SR, with a relative dielectric constant of ε_r of 3 and a thickness of 400 nm. Conventional photolithography is used to define the capacitors in the MIM stack. The top Au layer of the MIM stack is used as anode for the vertical diode. A 350 nm pentacene layer, the organic semiconductor, is evaporated through a shadowmask by HV-deposition. Finally, an aluminum (Al) cathode is evaporated through a second shadowmask.

The organic 64-bit transponder chip is made on a 25 µm thin plastic substrate using organic bottom-gate thin-film transistors. The organic electronics technology

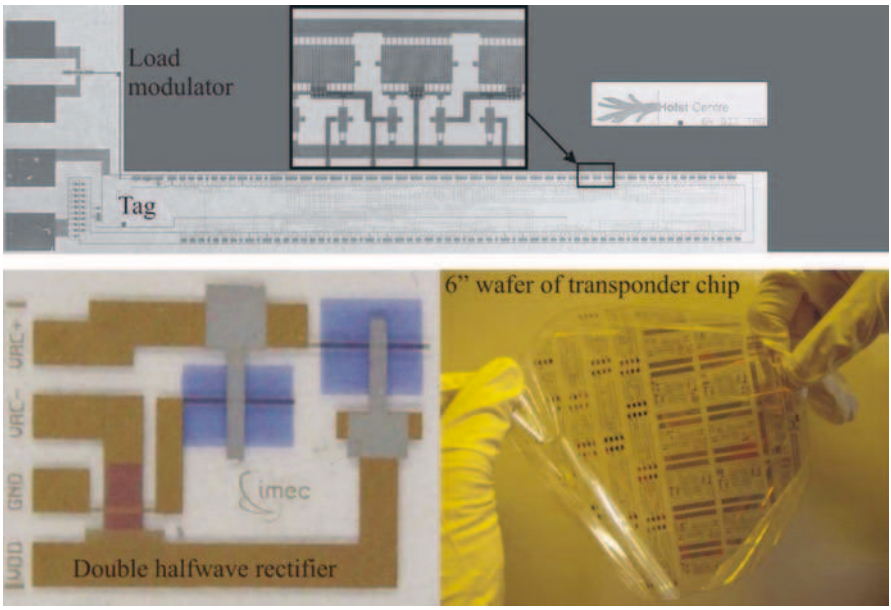


Fig. 2 Pictures of the transponder foil and the load modulator foil (*top*), the double half-wave rectifier foil (*bottom left*) and the 150 mm flexible wafer full of transponder chips (*bottom right*)

that is used, was developed by Polymer Vision for commercialization in rollable active matrix displays and is described elsewhere [6]. The insulator layers and the semiconductor layer are organic materials processed from solution. The transistors, with a typical channel length of 5 μm, exhibit an average saturation mobility of 0.15 cm^2/V s. A micrograph picture of the 64-bit transponder chip and the 150 mm wafer is also depicted in Fig. 2.

2.2 The RFID Measurement Setup

The complete tag is realized by properly interconnecting the contacts of the four foils, which we achieved in an experimental set-up where we plug the individual foils into sockets as shown in Fig. 3. Alternatively, we have also achieved tags by lamination of the foils, whereby electrically conductive glue is used to interconnect the different contacts of the individual foils.

The reader setup conforms to the ECMA-356 standard for "RF Interface Test Methods". It comprises a field generating antenna and two parallel sense coils (Fig. 3), which are matched to cancel the emitted field. By this method, only the signal sent by the RFID tag is read out at the reader side. The detected signal is then demodulated by a simple envelope detector (inset Fig. 9), being a diode followed by a capacitor and a resistor, and shown on an oscilloscope.

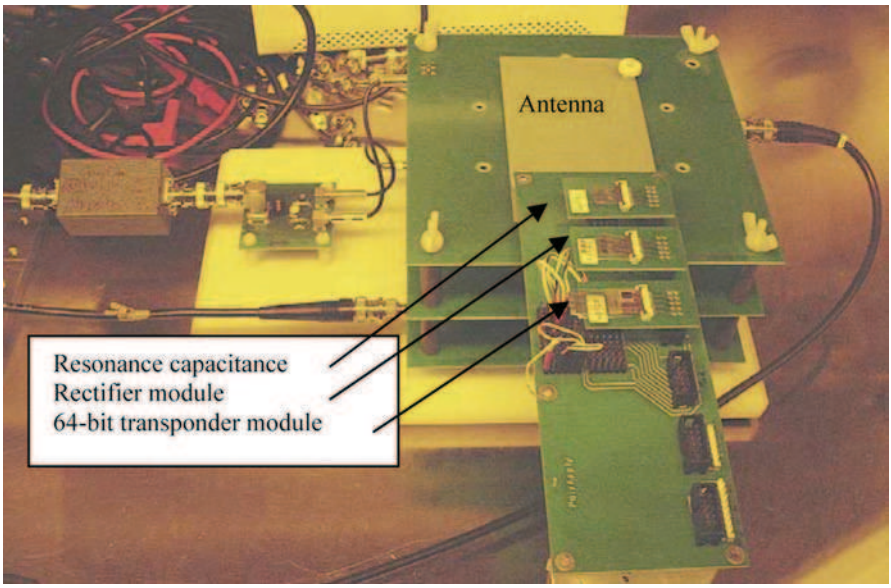


Fig. 3 Overview of the reader and RFID tag measurement setup. Foils are placed in sockets to ease manipulation

2.3 Organic Transponder Chip

Possible logic gate design architectures in unipolar, single threshold voltage logic are depletion-mode logic or enhancement-mode logic [1]. The gates in our organic RFID transponder chip are designed using the former, which is also known as zerovgs-logic. The choice for this type of logic was driven by the fact that the OTFTs used in this work show normally-on or depletion-mode behavior. The final design of the organic transponder chip has been made using only inverters and NAND-gates, both implemented in the zerovgs-logic. The gain of such an inverter at the trip point, for supply voltages of 10 and 20 V, is 1.75 and 2.25, respectively. Nineteen-stage ring oscillators of inverters operate at a frequency of 627 Hz when using 10 V supply voltage and at 692 Hz when using 20 V supply voltage.

The schematic of the transponder chip is depicted in Fig. 4. A 19-stage ring oscillator generates the clock signal when powered. This clock signal is used to clock the output register, the 3-bit binary counter and the 8-bit line select. The 8-bit line select has an internal 3-bit binary counter and a 3-to-8 decoder. This block selects a row of 8 bits in the code. The 3-bit binary counter drives the 8:1 multiplexer, selecting a column of 8 bits in the code matrix. The data bit at the crossing of the active row and column, is transported via an 8:1 multiplexer to the output register, which sends this bit on the rising edge of the clock to the modulation transistor. The 3 bits of the

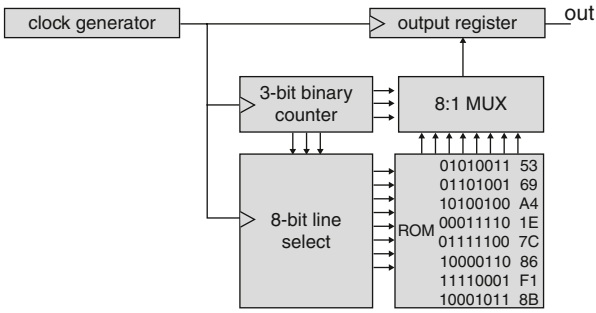


Fig. 4 Schematic overview of the digital logic part of the 64-bit transponder chip

3-bit binary counter are also used in the 8-bit line select block for selecting a new row after all 8 bits in a row are transmitted.

The 64-bit transponder foil comprises only 414 OTFTs. Figure 2 shows a micrograph image of the transponder foil. At 14 V supply voltage the 64-bit transponder foil generates the correct code at a data rate of 752 b/s, which is depicted in Fig. 5. Besides the 64-bit transponder foil, we also designed an 8-bit transponder foil. The main difference in the design is the complexity of the line select.

2.4 Organic Rectifier

The purpose of the rectifier in an RFID tag is to create a DC-voltage from the AC-voltage detected and generated by an antenna at the targeted base carrier frequency of 13.56 MHz. This frequency is selected because it is a standard in Si-based RFID tags, and will therefore enable partial compatibility with installed reader systems at 13.56 MHz. An important issue for organic RFID tags is the efficiency of the rectifier. A more efficient rectification generates the required DC voltage from a smaller AC input voltage. This implies larger reading distances for the RFID tags [3, 4].

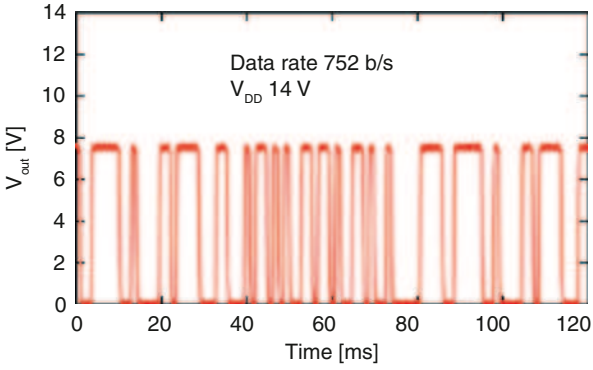


Fig. 5 Measured code of the 64-bit transponder chip at a supply voltage of 14 V

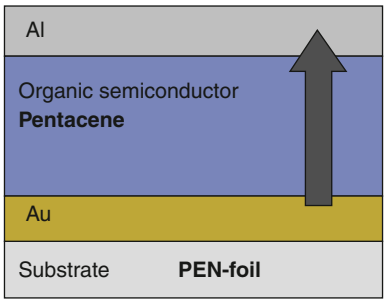


Fig. 6 Structure of a vertical, organic semiconductor-based Schottky diode

A rectifier comprises of diodes and capacitors. For organic diodes, two different topologies can be used: a vertical Schottky diode [7, 8] or a transistor with its gate shorted to its drain. The transistor with shorted gate-drain node is often considered as the most favourable topology because its process flow is equal to that used for the transistors in the digital circuit of the RFID tag. We have chosen to use the vertical diode structure because of its better intrinsic performance at higher frequencies compared to transistors as diodes [9].

The structure of a vertical, organic diode used to make the rectifier, is shown in Fig. 6. As depicted, we fabricated hole-only organic diodes with a layer of pentacene (as organic semiconductor) sandwiched between an Au- and Al-electrode on a 150 mm PEN foil. Because of their workfunctions, the Al-electrode blocks the injection of holes, whereas the Au-electrode permits the injection of holes.

We have fabricated a more efficient rectifier circuit, being a double half-wave rectifier. This rectifier comprises two diodes, each followed by a capacitor [3]. Figure 7a shows the schematic of this circuit. Both capacitors are 20 pF. The active area of the diodes is 500×200 µm.

A double half-wave rectifier circuit consists of two single half-wave rectifiers connected between the same nodes, with diodes connected as shown in Fig. 7a.

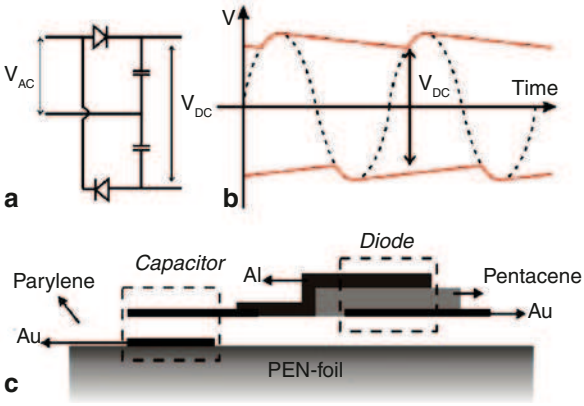


Fig. 7 **a** Schematic of a double half-wave rectifier, **b** schematic operation of a double half-wave rectifier, **c** vertical cross-section of the integrated capacitors and diodes on foil

Both single half-wave rectifiers rectify the AC input voltage: one rectifies the upper cycles of the AC input voltage, the other single half-wave rectifier rectifies the lower cycles of the input voltage. This is schematically depicted in Fig. 7b. The power and the ground voltage for the digital logic of the RFID tag are taken between both rectified signals (Fig. 7a, b). Therefore, a double half-wave rectifier yields about double the rectified voltage compared to a single half-wave rectifier.

2.5 Organic RFID Tag Using DC Load Modulation

The organic RFID transponder foil, the antenna coil, the HF-capacitor and the rectifier on foil are connected together to form an organic RFID tag. In DC load modulation mode, the modulation transistor (W/L = 5040 µm/5 µm) is placed behind the rectifier, as can be seen in Fig. 1. All foils are placed into sockets and connected as depicted in Fig. 3. The RFID reader is a 7.5 cm radius antenna which emits the field at a base carrier frequency of 13.56 MHz. In Fig. 8, the internal rectified voltage of this double half-wave rectifier in the organic RFID tag is plotted as a function of the field generated by the reader, for the tag antenna placed in the near-field of the reader antenna, at a distance of about 4 cm from the coil generating the readers 13.56 MHz RF field. As can be seen in the graph, 10 V rectified voltage is obtained at a 13.56 MHz electromagnetic field of about 0.9 A/m, and 14 V at 1.26 A/m. The latter is the voltage currently required by our 64-bit organic transponder chips. The ISO 14443 standard states that RFID tags should be operational at a minimum required RF magnetic field strength of 1.5 A/m.

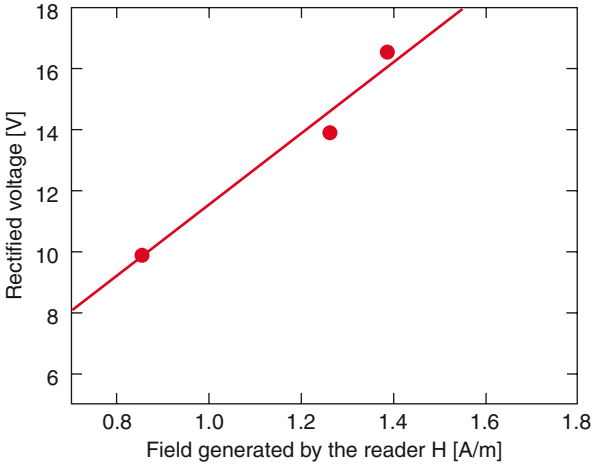


Fig. 8 Internal rectified voltage of a double half-wave rectifier generated in an organic RFID tag versus the 13.56 MHz magnetic field generated by the reader

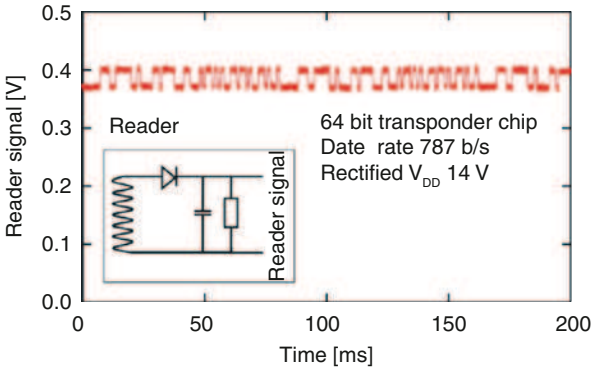


Fig. 9 Signal of the 64-bit RFID tag measured on the reader (unamplified reader signal). The envelope detector of the reader is depicted in the inset

The double half-wave rectifier circuit presented here therefore satisfies this ISO norm. After extrapolation of the measurement data, a DC voltage of 17.4 V can be obtained at a field of 1.5 A/m. If a single half-wave rectifier was used, the rectified voltage would be limited to 8–9 V, which is too low for current organic technology.

The obtained rectified 14 V drives the transponder chip, which sends the code to the modulation transistor. The signal sent from the fully integrated, plastic tag is received by the reader and subsequently visualized using a simple envelope detector (see inset Fig. 9) without amplification. The signal measured at the reader side is depicted in Fig. 9. This shows the fully functional, 64-bit RFID tag using an inductively-coupled 13.56 MHz RFID configuration with a data rate of 787 b/s. With a 0.7 V drop over the diode at the reader (envelope detector), a tag-generated signal of about 1.1 V is obtained, from which 30 mV is load modulation (modulation depth h = 1.4%).

Two of the reader standards at 13.56 MHz base carrier frequency are the proximity (ISO 14443) and vicinity readers (ISO 15693). The main difference between them is the coil radius, being 7.5 cm for the proximity reader and 55 cm for the vicinity reader. This results in a maximum readout distance of 10 cm for the proximity and 1 m for the vicinity reader. As mentioned earlier, the standard (ISO 14443) states also that the tag should be operational at an RF magnetic field of 1.5 A/m, which is significantly lower than the maximum allowed RF magnetic field of 7.5 A/m. One can calculate the required magnetic field at the antenna centre in order to obtain the required field to operate the tag. In our case, the required field for an 8-bit organic RFID tag was 0.97 A/m. This is depicted in Fig. 10. The dots in this graph show the experimental data at distances of 3.75, 8.75 and 13.75 cm with respect to the field generating antenna. This graph shows that it is possible to energize the 8-bit organic RFID tag at maximum readout distances for proximity readers below the maximum allowed RF magnetic field. The signal detected by the reader during the same experiment is depicted in Fig. 11 at distances of 5 and 10 cm with respect to the sense coil.

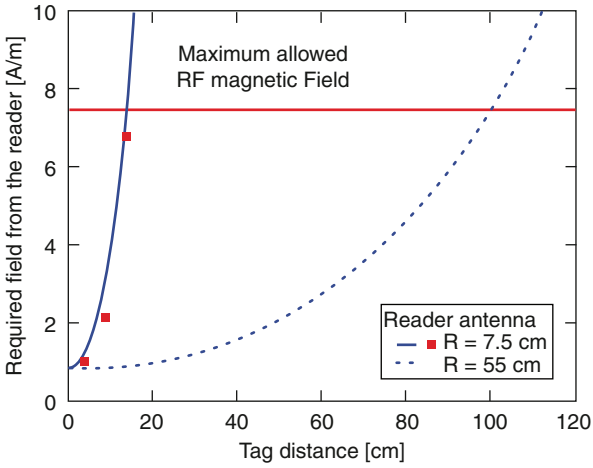


Fig. 10 Calculation and experimentally obtained data of the required RF magnetic field at the reader side as a function of the tag distance in order to generate the required RF magnetic field to operate the tag

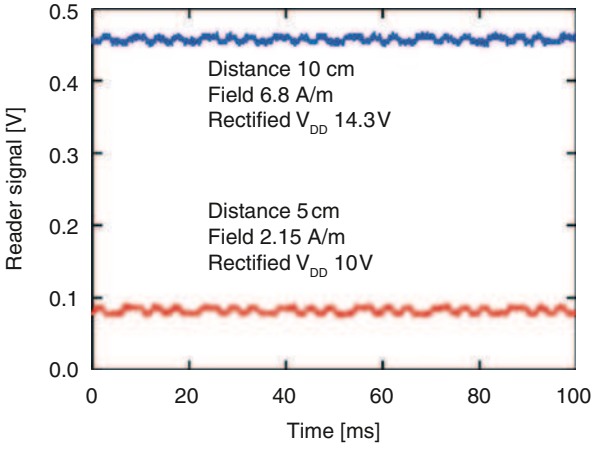


Fig. 11 Signal of the 8-bit RFID tag measured on the reader (unamplified reader signal) in DC load modulation mode at distances of 5 and 10 cm

3 Methods to Increase Robustness of Digital Circuits

In the realizations of organic RFID transponder chips discussed so-far [1, 2, 4], the most widespread technology is single-V_T p-type only. This technology has intrinsic limitations concerning integration of larger circuits, as a result of parameter variability. The move towards an organic complementary technology [10, 11] is the most favorable route to obtain robust organic circuitry [12] and first 4-bit tags in complementary organic technology [10] have been made. This requires, however, a matching n-type material and device performance. This has turned out to be difficult and complex route. Here, we present an alternative to improve the robustness of organic circuits [13]. It is based on p-type organic transistors equipped with a double gate [14–16]. The extra gate is used to vary the V_T, which effectively leads to a dual-V_T p-type technology. The dual-gate transistor configuration allows to in-

novate the topology of logic gates. We show inverter topologies with substantially increased gain and noise margin. Finally, we demonstrate a 64-bit organic RFID transponder chip based on these dual-gate Organic Thin-Film Transistors (OTFTs) for two different inverter topologies.

The availability of a second threshold voltage is beneficial for the characteristics of building blocks for digital circuits, similar to the silicon n-MOS logic used in the 1980s. One option to obtain multiple threshold voltages is by providing a second gate in the technology. The dual-gate organic TFT technology that is used in this work is described elsewhere [17]. The organic insulator layers and the p-type pentacene semiconductor are processed from solution. The transistors have a typical channel length of 5 µm and an average saturation mobility of 0.15 cm^2/V s. A cross-section image of the layers and a typical measured dual-gate OTFT are shown in Fig. 12. The second gate, or backgate, is coupled weaker to the channel than the gate, and therefore it is used as V_T-control gate. By varying the backgate between +30 and −30 V, the transistor threshold voltage can be controlled, leading from depletion-mode curve to more enhancement-mode curve, in agreement with earlier publications [18, 19].

Two different topologies for digital building blocks (e.g. inverters, NANDs) are the zerovgs-load logic and the diode-load logic. Both schemes are depicted in Fig. 13. Both inverters comprise a drive transistor, which is the pull-up transistor in p-type logic and a load transistor, i.e. the pull-down transistor. The ratio between the drive and load transistor for the diode-load logic is 10:1, however we have found that the zerovgs-load logic can be designed with a 1:1 ratio resulting in a significantly smaller chip area.

The transfer curves of zerovgs-load and diode-load inverters are depicted in Fig. 13. The V_T-control gate of the drive transistor can be used to move the trip point towards $V_{DD}/2$. At this point, in zerovgs-load inverters, the gain exceeds 11 and the noise margin is larger than 6 V at $V_{DD}=20$ V. Importantly, the typically low gain of diode-load inverters increases to 2, whereas the noise margin almost reaches 1.5 V.

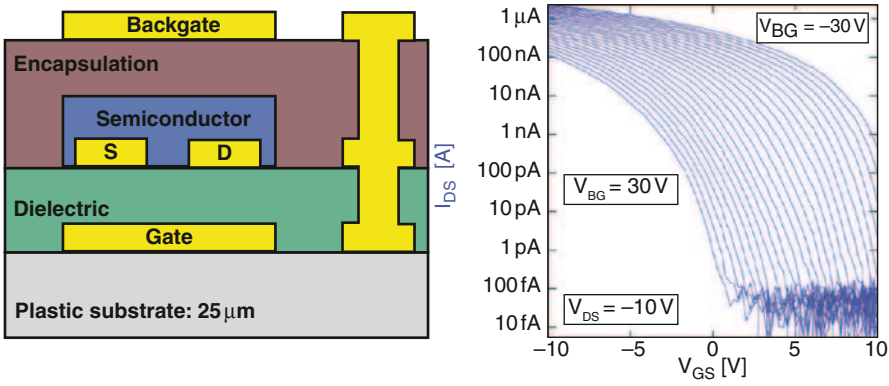


Fig. 12 Cross-section of the dual-gate OTFT technology (*left*) and a typical measured transfer characteristic of this OTFT when using the backgate as V_T-control gate (*right*). The channel width equals 140 µm, the channel length 5 µm

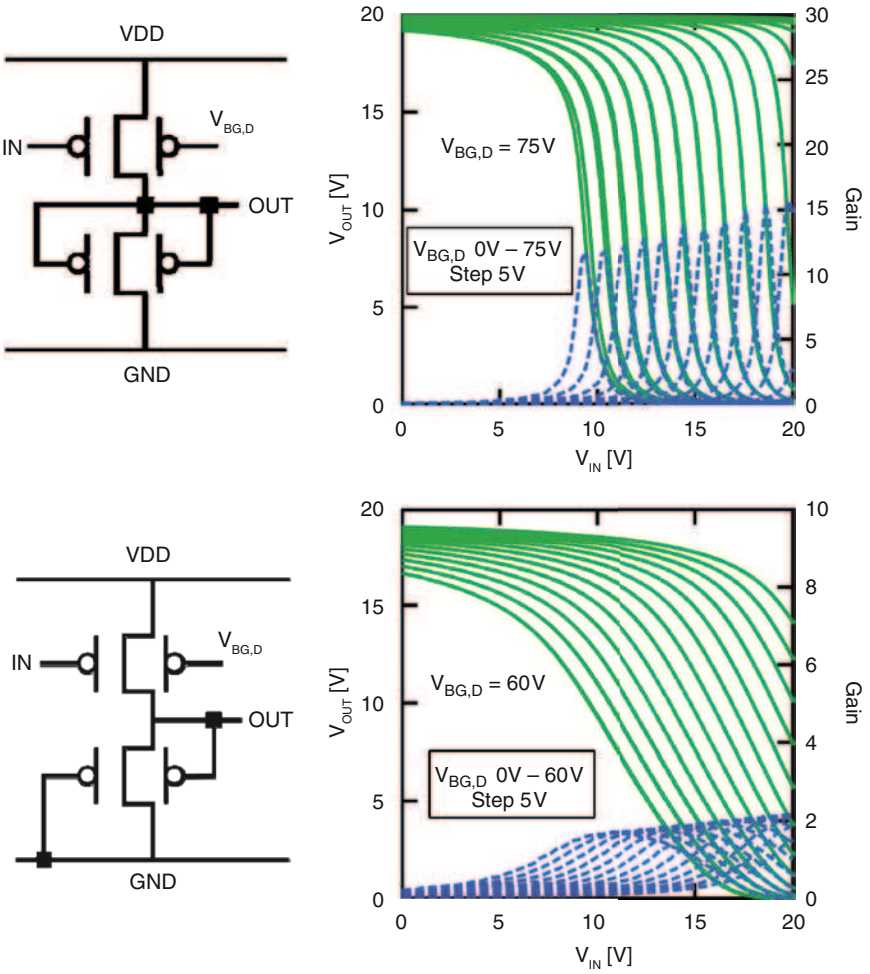


Fig. 13 Architecture of inverter topologies in dual gate technology for (*top*) zerovgs-load inverter and (*bottom*) diode-load inverter. Transfer curves of these inverters are plotted to the right for a supply voltage of 20 V, as a function of the voltage on the V_T-control gate of the drive transistor

99-Stage dual-gate ring oscillators with zerovgs-load and diode-load architectures were fabricated and characterized. In Fig. 14, we show the extracted stage delays of NAND-gates and inverters, and compare them to reference single-gate zerovgs-load gates fabricated on the same foil. The fastest family is the diode-load topology, with a stage delay of 2.27 µs. The zerovgs-load topology is an order of magnitude slower (τ_{stage} = 26 µs), due to the limiting pull-down current and the high parasitic output-capacitance. This capacitance is also present in the single-gate depletion-load architecture but the drive-transistor current is now substantially higher, resulting in a smaller stage delay for the single gate variant.

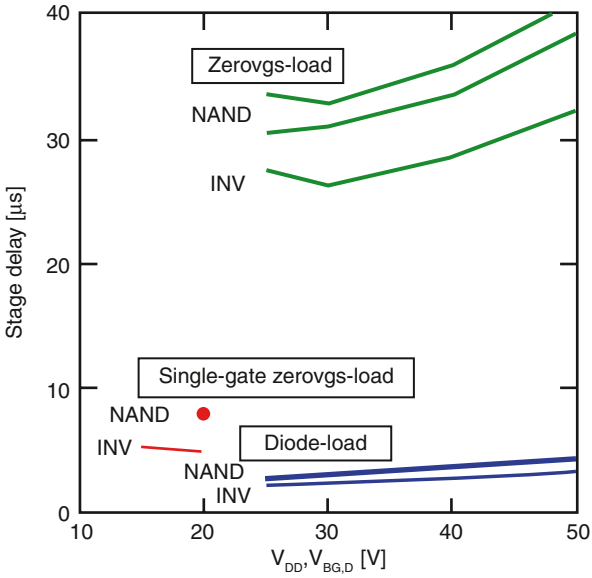


Fig. 14 The stage delay is plotted as a function of the V_T-control voltage of the drive transistor for dual-gate zerovgs-load and diode-load inverters for a supply voltage of 20 V. The stage delay of a single-gate zerovgs-load inverter is shown for reference, also for a supply voltage of 20 V

3.1 Organic RFID Transponder Chip with Dual-Gate Architecture

Next, we fabricated 64-bit RFID transponder chips similar to earlier designs (Sect. 2.3.2), but with a clock from a 33-stage ring oscillator to prevent racing of the signal. For the diode-load logic, the ratio between drive and load transistor is 10:1 leading to a transponder chip area of 74.48 mm^2. Figure 15 shows the output signal of the chip designed in diode-load topology for V_{DD}=20 V and $V_{BG,D}$=45 V. The data rate is 4.3 kb/s, more than double the fastest single gate transponder chips

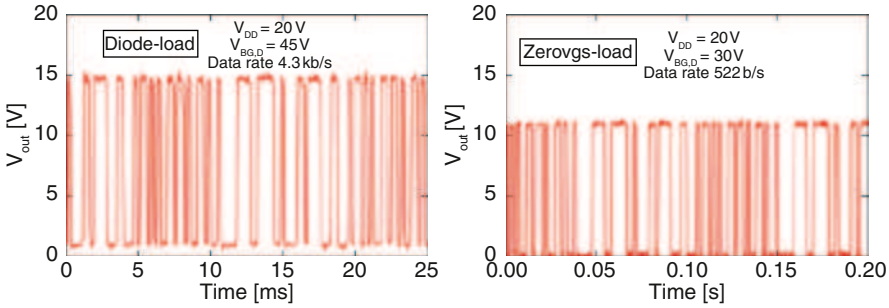


Fig. 15 The output signal of a 64-bit organic RFID transponder chip with (*left*) diode-load configuration for a supply voltage of 20 V and a backgate voltage of 45 V and (*right*) with zerovgs-load configuration for a supply voltage of 20 V and a backgate voltage of 30 V. The corresponding data rates are respectively 4.3 kb/s and 522 b/s

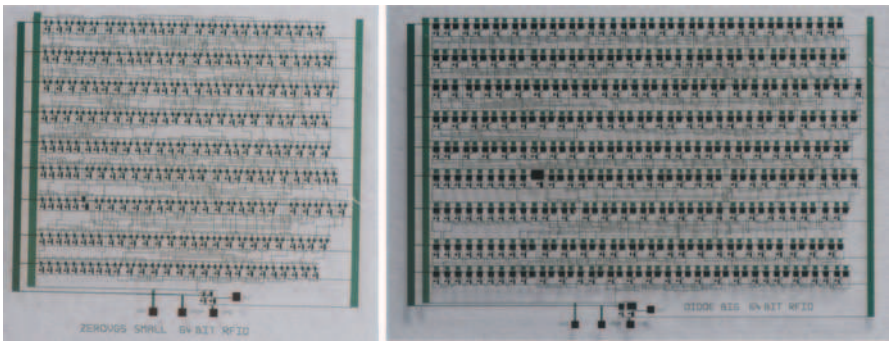


Fig. 16 Photograph of the 64-bit organic RFID transponder chip designed with zerovgs-load configuration (*left*) and diode-load configuration (*right*)

shown in [4]. We verified on three different foils that some chips start operating at V_{DD} as low as 10 V and all chips work at 15 V.

The 64-bit transponder chip in zerovgs-load topology is only 45.38 mm^2 in size, thanks to this logic having a 1:1 ratio. In Fig. 16, the output is shown for $V_{DD}=20$ V and $V_{BG,D}=30$ V. Also this transponder chip has been measured on three different foils. All measured transponder chips are operational at $V_{DD}=10$ V. The data rate at $V_{DD}=20$ V is 522 b/s. These findings are fully in line with the fact that the zerovgs topology has higher noise margin but slower stage delay than the diode-load topology. Figure 16 shows a photograph of both 64-bit transponder chips.

4 Vision for Commercialization

RFID is an important technology in logistics, retail, automation, anti-counterfeit protection, identification and various other fields. It allows for transmission of data (an identification code in the memory of the transponder chip) via radio waves from a transponder to a reader without the need for line of sight. In standard, silicon based electronics, this is a large market, segmented in several areas of applications. Several radio communication frequencies are used, spanning from tens of kHz to about 1 GHz.

In the last years, a protocol named Electronic Product Code "EPC" [20] has been developed for wireless identification in high-volume logistics applications like retail. It is widely used already today e.g. on pallet level logistics. The next step is to use EPC tags for the package level and the longer term target will be its use for individual items ("item-level tagging"). Such tags will likely be further equipped with sensors that can capture information of the environment (gases, temperature, …) and transmit that information along with the identity of the object. This leads to a technology for smart labels, providing smartness to objects.

One of the major limits of the EPC tags on their way towards item-level tagging is the price of the single silicon transponder chip on the tag. The costs of the tags (or "inlays") are dominated by the cost of the silicon chip, followed by the cost of the antenna and the assembly costs. Today, standard tags in HF cost in the order of 0.15–0.20 US$/inlay. Standard UHF tags are less expensive, but still in the order of 0.10–0.15 US$/inlay. The major cost difference between UHF and HF tags is related to the antenna, which is more expensive in the HF range. For low-cost products, the cost of the silicon transponder chip is significantly too high.

With organic RFID tags, the price of RFID tags is expected to go down significantly compared to standard tags. The reason is essentially that the transponder chip made with organic technology can be significantly lower in cost than with silicon. Indeed, organic technologies are applied at low temperature (little energy), directly on low-cost substrates (saving packaging costs) and at high processing speeds (resulting in lower cost-of-ownership of fabs).

Organic-based technologies furthermore confer superior mechanical properties to the assembled tag. Conventionally, the lamination of the RFID "inlay" in cardboard, paper or thin plastic results in an end-product with uneven topology, due to inlay thickness variations mainly caused by the presence of the rigid silicon chip—and this limits the possibilities of roll-to-roll handling and printing. Furthermore, the RFID inlay with brittle Si chips and connections are reputed for poor robustness, in particular during roll-to-roll handling. Thinned-down silicon chips exist, but they are even more expensive and more fragile and that limits their ultimate integration speed into inlays. In contrast, organic RFID chips are thin and flexible, which solves both the handling and the yield issues. Finally, by their large-area nature, organic technologies are more adequate to later integrate sensors onto RFID tags to create smart labels. That addition will lead to diversified product portfolio in markets beyond RFID.

The challenge is to achieve organic tags with maximum adherence to the EPC protocol. Adherence to the standard protocol is important for widespread acceptance. Ideally, the tags should become compatible with installed readers.

Some of those EPC specifications have already been met by plastic tags in recent years, namely the transmission of 64-bit and 96-bit codes (with supplemental bits for redundancy and destroy code) [1, 2, 4], as well as the use of HF (13.56 MHz) base carrier frequency compatible with regulations concerning human exposure to electromagnetic fields [4] and basic anti-collision protocols [4]. In the next section, we will describe recent evolutions towards higher bitrates, getting close to the requirement of the EPC specifications (i.e., up to 52.969 kb/s [20]).

5 A 50 kHz RFID Transponder Chip

The critical factor to address the circuit speed is the current drive of the transistors, which is determined by (i) the carrier mobility, (ii) the specific capacitance of the gate dielectric, and (iii) the inverse of the channel length. Vapor-deposited

pentacene [21] has a mobility, in our bottom-contact devices, of 0.5 cm^2/V s. The isolation of the transistors is achieved by an integrated shadow mask [22], shown in Fig. 17, that results in a reliable isolation of the semiconductor area and off-currents below 10 pA. As gate dielectric, we use sputtered Al_2O_3, with a specific capacitance of 70 nF/cm^2, treated by a self-assembled monolayer of Trichloro(phenethyl) silane. That, in turn, allows for some downscaling the transistor channel length, in our case to 2 µm, while maintaining a reasonable output resistance in saturation of Ro = 9.75 MΩ at 0 V gate-source voltage. A cross-section of this process flow is depicted in Fig. 17.

This organic thin-film circuit technology allows to design with smaller overlap capacitance and to downscale the transistor channel length, within the boundaries achievable by existing high-throughput tools (e.g. steppers used in backplane manufacturing). Specifically, we investigated downscaling of the channel lengths (L) of the circuits from 20 and 2 µm. We also limited the gate-source and gate-drain overlap capacitances by decreasing the width of the finger-shaped source and drain contacts (that overlap the gate) from 5 to 2 µm. All devices and circuits are processed on 25 µm thick polyethylene naphthalate (PEN) foil, which was, during processing, laminated on a 150 mm carrier substrate, and delaminated after completion of the process.

Typical transfer and output curves of transistors fabricated in this technology, having L = 5 and 2 µm, are depicted in Fig. 18. The transistors are normally-on and their charge carrier (hole) mobility exceeds 0.5 cm^2/V s. The output resistance in saturation at gate-source voltage of 0 V is 207 ± 8 MΩ for L = 5 µm and 9.8 ± 0.2 MΩ for L = 2 µm.

Figure 19 depicts the inverter stage delay as a function of supply voltage, as extracted from 19-stage ring oscillators, for transistors with channel lengths varying from 20 to 2 µm and with gate-overlap of the transistor-fingers ranging from 5 to

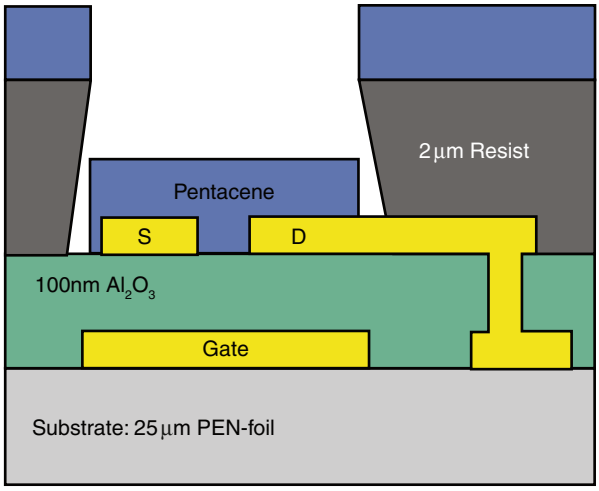


Fig. 17 Cross-section of the organic semiconductor-based thin-film transistor process on foil having Al_2O_3, treated with a self-assembled monolayer of Trichloro(phenethyl) silane, as gate dielectric

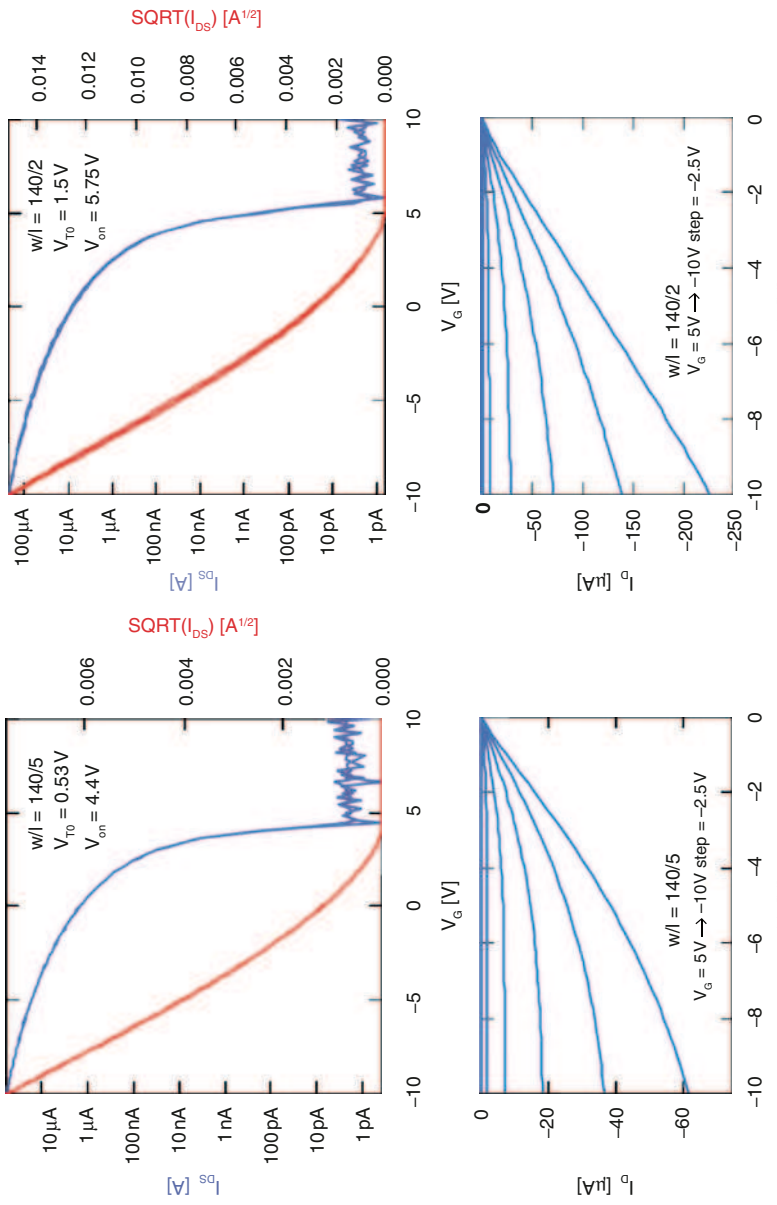


Fig. 18 Typical measured (*top*) transfer and (*bottom*) output curves of transistors with (*left*) W/L = 140/5 and (*right*) W/L = 140/2. The mobility exceeds 0.5 cm^2/V s, off-current is below 10 pA

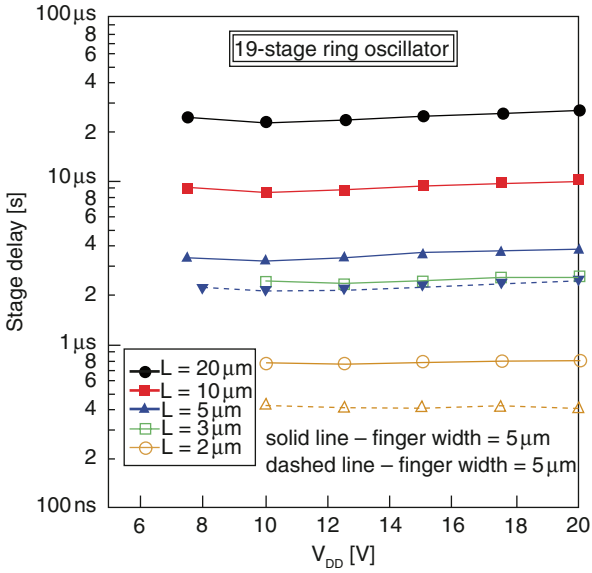


Fig. 19 Overview of stage delays versus supply voltage, measured on 19-stage ring oscillators with varying channel lengths, between 20 and 2 µm (*color and symbol codes*) and source/drain finger sizes of 5 µm (*solid lines*) and 2 µm (*dashed lines*)

2 µm. Stage delays below 1 µs, and as low as 400 ns, are shown at V_{DD} as low as 10 V. The effect of decreasing the overlap capacitance is also shown in Fig. 19 for the circuits having a channel length of 2 and 5 µm: shrinking the overlap from 5 µm (solid lines) to 2 µm (dashed lines) improves the stage delay by a factor of 1.5–2.

We proceeded with the design and realization of 8-bit RFID transponder chips, having a channel length of 2 µm and either 5 or 2 µm finger widths. Figure 20 shows the photographs of the 150 mm wafer and a zoom of a single die. Figure 21

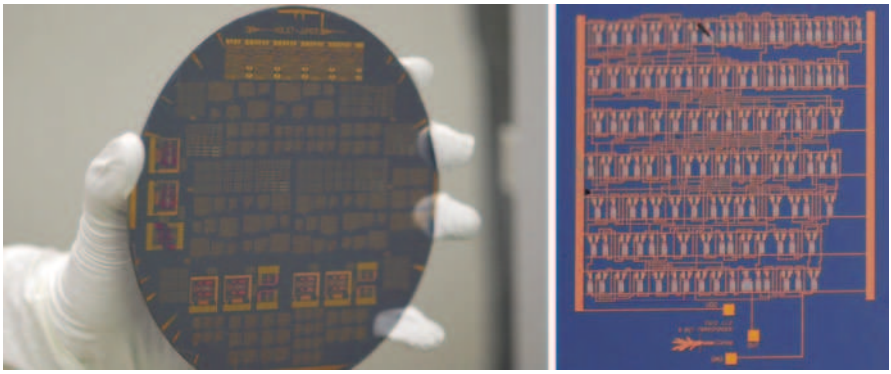


Fig. 20 Photograph of the 150 mm foil on a carrier wafer comprising all measured circuits (*left*). Photograph of the 8-bit RFID transponder chip on foil (*right*). The die size is 24.73 mm^2. The design comprises 294 transistors

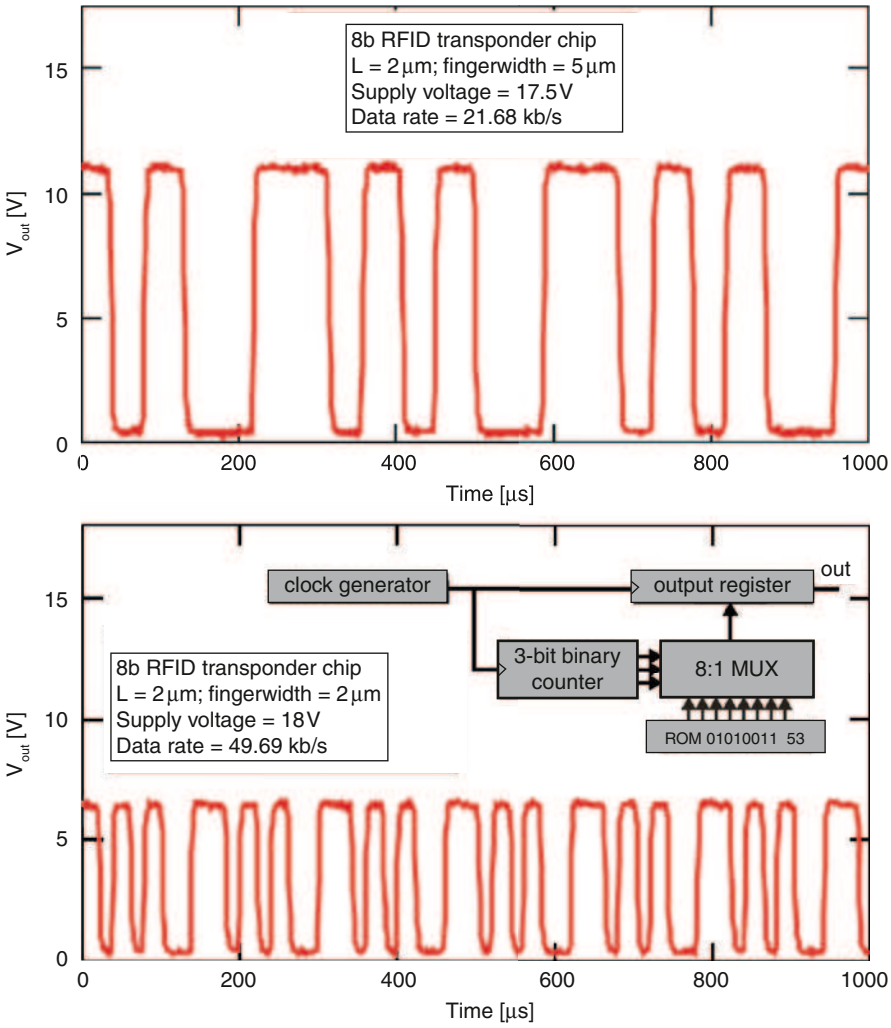


Fig. 21 Measured signal of the 8-bit RFID transponder chip having transistors with channel length of 2 µm: (*top*) width of source-drain fingers of all transistors is 5 m; (*bottom*) width of source-drain fingers of all transistors is 2 µm. The supply voltage for the 8-bit transponder chips were (*top*) 17.5 V and (*bottom*) 18 V, and data rates are respectively 21.68 and 49.69 kb/s. The inset of the bottom picture shows the schematic overview of the digital logic portion of the 8-bit transponder chip

depicts the output signal of both types of transponders. In agreement with the two-fold faster inverter stage delay for the 2 µm fingers, the data rate of this transponder is also twice as high as that of the design with 5 µm fingers. The obtained data rate of the 8-bit transponder with channel length and fingers of 2 µm reaches 50 kb/s [5].

6 Conclusions

In this chapter, we presented the technology, designs and implementation of an inductively-coupled passive 64-bit organic RFID tag which is fully functional at a 13.56 MHz magnetic field strength of 1.26 A/m. This RF magnetic field strength is below the minimum required RF magnetic field stated in the ISO standards. The 64-bit transponder chip employs 414 OTFTs and yields a data rate of 787 b/s. Also an 8-bit transponder chip was measured in DC load modulation configuration and could be readout at a distance of 10 cm, which is the expected readout distance for proximity readers.

We also demonstrated more robust transponder chips by use of a dual-gate technology, whereby the backgate controls the threshold voltage of the transistor. This yielded data rates of 4.3 kb/s for 64-bit transponder chips with a dual-gate diode-load topology. The supply voltage could be lowered to 10 V when using a dual-gate zerovgs-load implementation.

Finally, we elaborated on an 8-bit transponder chip having data rates that are EPC-compatible. This has been achieved in a thin-film transistor technology by using a high-k Al_2O_3 gate dielectric, by scaling the channel lengths down to 2 μm and by reducing the overlap capacitance of the parasitic source-gate and drain-gate capacitors.

References

1. E. Cantatore, T.C.T. Geuns, G.H. Gelinck, E. van Veenendaal, A.F.A. Gruijthuijsen, L. Schrijnemakers, S. Drews, D.M. de Leeuw, A 13.56-MHz RFID system based on organic transponders. IEEE JSSC. **42**(1), 84–92 (2007)
2. A. Ullmann, M. Böhm, J. Krumm, W. Fix, Polymer Multi-Bit RFID transponder, in *International Conference on Organic Electronics (ICOE)*, Abstract 53, Eindhoven, June 4–7 2007
3. K. Myny, S. Steudel, P. Vicca, J. Genoe, P. Heremans, An integrated double half-wave organic Schottky diode rectifier on foil operating at 13.56 MHz. Appl. Phys. Lett. **93**, 093305 (2008)
4. K. Myny, S. Steudel, P. Vicca, M.J. Beenhakkers, N.A.J.M. van Aerle, G.H. Gelinck, J. Genoe, W. Dehaene, P. Heremans, Plastic circuits and tags for 13.56 MHz radio-frequency communication, Solid State Electron. **53**(12), 1220–1226 (2009)
5. K. Myny, S. Steudel, S. Smout, P. Vicca, F. Furthner, B. van der Putten, A.K. Tripathi, G.H. Gelinck, J. Genoe, W. Dehaene, P. Heremans, Organic RFID transponder chip with data rate compatible with electronic product coding, Organic Electronics **11**(7), 1176–1179 (2010) (available online)
6. H.E.A. Huitema, Rollable displays: the start of a new mobile device generation, in *7th Annual Flexible Electronics & Displays Conference USDC*, Phoenix, Arizona, January 2008
7. S. Steudel, K. Myny, V. Arkhipov, C. Deibel, S. De Vusser, J. Genoe, P. Heremans, 50 MHz rectifier based on an organic diode. Nat. Mater. **4**, 597–600 (2005)
8. B.N. Pal, J. Sun, B.J. Jung, E. Choi, A.G. Andreou, H.E. Katz, Pentacene-zinc oxide vertical diode with compatible grains and 15-MHz rectification. Adv. Mater. **20**, 1023–1028 (2008)
9. S. Steudel, S. De Vusser, K. Myny, M. Lenes, J. Genoe, P. Heremans, Comparison of organic diode structures regarding high-frequency rectification behavior in radio-frequency identification tags. J. Appl. Phys. **99**, 114519 (2006)

10. R. Blache, J. Krumm, W. Fix, Organic CMOS circuits for RFID applications, in *IEEE ISSCC Dig. Tech. Papers*, (February 2009), pp. 208, 209
11. K. Ishida, N. Masunaga, Z. Zhou, T. Yasufuku, T. Sekitani, U. Zschieschang, H. Klauk, M. Takamiya, T. Someya, T. Sakurai, A stretchable EMI measurement sheet with 8×8 coil array, 2 V organic CMOS decoder, and −70 dBm EMI detection circuits in 0.18 μm CMOS vol. 473, in *IEEE ISSCC Dig. Tech. Papers*, (February 2009), pp. 472, 473
12. D. Bode, C. Rolin, S. Schols, M. Debucquoy, S. Steudel, G.H. Gelinck, J. Genoe, P. Heremans, Noise margin analysis for organic thin film complementary technology. IEEE Trans. Electron. Dev. **57**(1), 201–208 (2010)
13. K. Myny, M.J. Beenhakkers, N.A.J.M. van Aerle, G.H. Gelinck, J. Genoe, W. Dehaene, P. Heremans, Robust digital design in organic electronics by dual-gate technology, in *IEEE ISSCC Dig. Tech. Papers*, (February 2010), pp. 140, 141
14. J.B. Koo, C.B. Koo, C.H. Ku, J.W. Lim, S.H. Kim, Novel organic inverters with dual-gate pentacene thin-film transistor. Org. Electron. **8**, 552–558 (2007)
15. M. Spijkman, E.C.P. Smits, P.W.M. Blom, D.M. de Leeuw, Y. Bon Saint Côme, S. Setayesh, E. Cantatore, Increasing the noise margin in organic circuits using dual gate field-effect transistors. Appl. Phys. Lett. **92**, 143304 (2008)
16. M. Takamiya, T. Sekitani, Y. Kato, H. Kawaguchi, T. Someya, T. Sakurai, An organic FET SRAM for braille sheet display with back gate to increase static noise margin, in *IEEE ISSCC Dig. Tech. Papers*, (February 2006), pp. 276, 277
17. G.H. Gelinck, H.E.A. Huitema, M. van Mil, E. van Veenendaal, P.J.G. van Lieshout, F. Touwslager, S.F. Party, S. Sohn, T.H. Whitesides, M.D. McCreary, A rollable, organic QVGA display with field-shielded pixel architecture. J. SID (Soc. Information Display). **14**, 113–118 (2006)
18. G.H. Gelinck, E. van Veenendaal, R. Coehoorn, Dual-gate organic thin-film transistors. Appl. Phys. Lett. **87**, 073508 (2005)
19. S. Iba, T. Sekitani, Y. Kato, T. Someya, H. Kawaguchi, M. Takamiya, T. Sakurai, S. Takagi, Control of threshold voltage of organic field-effect transistors with double-gate structures. Appl. Phys. Lett. **87**, 023509 (2005)
20. EPC standard, http://www.epcglobalinc.org/standards/specs/
21. C. Rolin, S. Steudel, P. Vicca, J. Genoe, P. Heremans, Functional pentacene thin films grown by in-line organic vapor phase speeds above 2 m/min. Appl. Phys. Exp. **2**, 086503 (2009)
22. S. De Vusser, S. Steudel, K. Myny, J. Genoe, P. Heremans, Integrated shadow mask method for patterning small molecule organic semiconductors. Appl. Phys. Lett. **88**, 103501 (2006)